Università degli Studi di Bergamo
School of Engineering
Department of Engineering and Applied Sciences
Doctoral Programme In Engineering and Applied Sciences
Doctoral curriculum CD4 - Mathematics and Applied Statistics

Bilevel programming models for power generation capacity expansion planning with revenue adequacy constraints

Doctoral Dissertation of:
Martina Gherardi

Supervisor:
Prof. Maria Teresa Vespucci

MBS Consulting tutor:
Ing. Virginia Canazza

The Chair of the Doctoral Program:
Prof. Alessandra Marini

A.A. 2022/2023 – XXXVI Cycle

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, Maria Teresa Vespucci, for her guidance and belief in my potential from the beginning. Additionally, I would like to extend my thanks to Giovanni Micheli for his availability, and to Renata Mansini and Giovanni Felici for their valuable comments in reviewing this work.

At MBS Consulting, special thanks go to my tutor, Virginia Canazza, and to Giacomo Ciapponi and Federico Clementi for their significant contributions in introducing me to the electricity market and applying the modeling framework to a real test case.

I am deeply grateful to Jalal Kazempour for his availability and time, for his ability to introduce me to some of the ideas explored in this thesis, and for welcoming me into his amazing EMA group during my visiting period in Copenhagen. The friendly atmosphere and good times shared with the group will always be in my heart. I would like to say a big thank you to the lovely Andrea. I really miss our office chats!

An "unbounded" (or mega?) thank you to Fabrizio Lacalandra for his invaluable mentorship, inspiration, and willingness to share his passion, knowledge, and curiosity in both mathematical approaches and policy/regulatory aspects during each of the countless teams.

To my friend Federica, thank you for your closeness and unwavering understanding. I look forward to meeting the beautiful Isabel, whose birth coincides exactly with the completion of this thesis.

My heartfelt thanks to Nicolas, for the big dreams (and the handmade gifts) and our growth together. I will always be immensely grateful to him, as well as to Giacomo and Debora, for welcoming me as part of their family throughout my PhD journey.

Finally, a huge thank you to my "gigantic" family. To my mother Roberta, for teaching me that the combination of ability and hard work pays off in the long run. To my father Carmelo, for teaching me the value of optimism, starting with a smile even if things do not go as desired. To Cristina, my example of independence and ability to reinvent oneself. To Giulia, my example of leaving a lasting impression. To Pietro, my example of constancy and tirelessness. To Francesca, my brilliant study partner despite the 9-year age difference. As your older sister, I am here to be a reference point for you. In conclusion, I would like to thank my beloved grandparents Guido and Clementina, who always served as second parents, and my grandmother Ida, who watches over me from above.

ABSTRACT

All over the world electrical systems are undergoing unprecedented evolution driven by technological advancements, environmental considerations, and changing energy consumption patterns. Several trends are expected to shape the future of power systems. In this context, quantitative models, such as optimization, simulations, and machine learning, will play more and more a crucial role in addressing the associated challenges both in short-term operation and in long-term planning.

In this thesis, we have focused preeminently on an innovative modeling development for the *Generation Expansion Planning* (GEP) problem, customized to address the forthcoming demand for decision support in the Italian electricity market. The GEP is formulated as a bilevel optimization problem, where *Revenue Adequacy* (RA) is ensured for different technologies by considering a detailed definition of market revenues with meaningful market clearing prices. This includes zonal distinctions as perceived by the Italian electricity market to take into account price signals. In addition, the model incorporates side payments that can serve as indicators for Capacity Remuneration Mechanism (CRM)-like auction. To ensure the realistic operation of the model, we have suggested two novel sets of constraints. These are designed to prevent withholding strategies and to deal with price indeterminacy.

The problem's bilevel nature, combined with revenue modeling, results in bilinear components. In addition to examining state-of-the-art linearization techniques, we have developed an exact linearization method for the product of two continuous variables, namely prices and quantities. To address the difficulty of finding feasible solutions, we have defined warm-start algorithms. Furthermore, we have proposed an innovative approach that utilizes "hybrid" modeling of the complementarity slackness conditions. This advancement has proven to be successful in narrowing the gap and proving optimality.

The thesis is organized as follows.

In Chapter 1 we introduce the GEP problem and provide a background on the electricity markets together with a central problem about the *missing money* and *revenue adequacy* in electricity markets. Both concepts will be a *fil rouge* through the entire thesis. In Chapter 2 we deepen the literature review on both the pricing scheme (2.1) in several markets and on the specific proposals for the GEP (2.2). For the first theme, we discuss important issues about non-convexity in pricing schemes while for the GEP we distinguish between traditional GEP and those that account for the critical aspect of the

revenue adequacy. In Chapter 3 we then switch to a brief exemplification of Mathematical Programming and to a presentation of a general taxonomy for optimization problems. Subsequently, we focus on a class of optimization problems that will be used for the proposed models, the bilevel optimization ones with their hierarchy structure. Notably, in (3.3) we discuss reformulations approaches.

In Chapter 4 we step into the models for the Market Operator and propose a modification of the Italian one w.r.t the inclusion of the technical minimum for certain generation plants. We also discuss important issues related to price indeterminacy in the Market Operator problem and strategic withholding of capacity by generators in our specific setting. Then a discussion about how to model the Objective Function of the GEP problem is analyzed and two main high-level models for the GEP are proposed. In Chapter 5 we then detail the two proposed models, named GEP-RA (1) and GEP-RA (2), with all the blocks of constraints that are therein discussed in formal details. Notably, we present an innovative formulation for a specific problem aiming at avoiding withholding strategies. Additionally, for the peculiar issue of indeterminacy of the clearing price, we propose another innovative block of constraints, that we later reformulate in (5.3) by means of an extended formulation. We conclude this chapter by giving a mixed-binary version of the proposed models in (5.4). In Chapter 6 we take the proposed formulations and derive reformulations of them in order to be solved by general purposes optimization solvers. Notably, we explicate the optimality conditions of the lower level of each bilevel problem and give single-level equivalent models. Then in (6.2) we propose a simple, yet effective, linearization of relevant bilinear components in the objective function(s) and in some constraints. In Chapter 7 we propose different approaches aimed at enhancing the computational efficiency. We tackle the challenge of identifying feasible solutions by introducing specialized cuts and developing auxiliary problems that effectively warm-start the GEP-RA (2) model. To address the task of proving optimality, we innovatively combine different formulations of the complementarity slackness conditions.

In Chapter 8 we are ready to give a description of the actual software implementation realized, and of the Italian test case that has been constructed and that will be used for the computational tests. This test case, together with a simple synthetic one, is then used for the comparison of the different proposed models both from a conceptual point of view and from the point of view of resolution performance. A set of disaggregated and aggregated output is presented and the results of the simulations in terms of modeling and performances are discussed. In Chapter 9 we describe future research directions in terms of modeling choices, i.e. including the transmission network and uncertainty, and others related to more efficient methodologies for solving the proposed problems that remain of very high complexity, e.g. formulations and decomposition approaches.

In Chapter 10 we end the journey by recalling the main findings and drawing conclusions of the present thesis.

Table of Contents

Acronym & Notation

ACER Agency for the Cooperation of Energy Regulators

ADMM Alternating Direction Method of Multipliers

AIC Average Incremental Cost

ARERA Autorità di Regolazione per Energia Reti e Ambiente (Italian Regulatory Authority for Energy, Networks and Environment)

CCGT Combined Cycle Gas Turbine

CfD Contracts for Difference

CH Convex Hull

CHP Convex Hull Price

CHPri Convex Hull Primal

CLARA Clustering LARge Applications

CP Central Planner

CRM Capacity Remuneration Mechanisms

CSC Complementarity Slackness Conditions

DAM Day Ahead Market

DPA Dual Pricing Algorithm

ELMP Extended Locational Marginal Prices

ENP Energy Not Provided

ES Energy Storage

GEP Generation Expansion Planning

GEP-RA Generation Expansion Planning with Revenue Adequacy

GTEP Generation and Transmission Expansion Planning

GU Generalized Uplift

IP Integer Programming

IP+ Integer Programming Positive

KKTs Karush-Kuhn-Tucker optimality conditions

LC Limited Compensation

LCOE Levelized Cost of Energy

LD Loose Dispatchable

LL Lower-Level

LOC Lost Opportunity Costs

LP Linear Program

MACSE Meccanismo di Approvvigionamento di Capacità di Stoccaggio Elettrico (Italian mechanism for the acquisition of storage capacity)

MASE Ministero dell'Ambiente e della Sicurezza Energetica (Italian Ministry of Environment and Energy Security)

MB Mercato di Bilanciamento (Italian Balancing Market)

MGP Mercato del Giorno Prima (Italian DAM)

MI Mercato Infragiornaliero (Italian Intraday Market)

MIBP Mixed-Integer Bilinear Program

MILP or MIP Mixed-Integer Linear Program

MINLP Mixed-Integer Nonlinear Programming

MIQCP Mixed-Integer Quadratically Constrained Program

MO Market Operator

modIP Modified Integer Programming

MP Mathematical Programming

MPCC Mathematical Program with Complementarity Constraints

MSD Mercato per il Servizio di Dispacciamento (Italian Ancillary Services Market)

MTE Mercato a Termine (Italian Forward Market)

MWP Make-Whole Payments

NEMO Nominated Energy Market Operator (Gestore dei Mercati Energetici, GME in Italy)

NLP Nonlinear Program

NRAs National Regulatory Authorities

OCGT Open Cycle Gas Turbine

OF Objective Function

P-D Primal Dual

PD Partial Dispatchable

PM Profit Maximization

PNIEC Piano Nazionale Integrato Energia e Clima (Italian National Energy and Climate Plan (NECP))

PUN Prezzo Unico Nazionale (Italian single price)

R Restricted

RA Revenue Adequacy

RAPA Revenue Adequate Price Adders

REMIT Regulation on wholesale Energy Market Integrity and Transparency

RES Renewable Energy Sources

RNP Reserve Not Provided

RQmin Relaxed Minimum Operation

SLR Semi-Lagrangian Relaxation

SOS1 Special Order Sets of type 1

SOS2 Special Order Sets of type 2

TSO Transmission System Operator (Terna in Italy)

UC Unit Commitment

UCED Unit Commitment and Economic Dispatch

UL Upper-Level

VoLL Value of Lost Load

Sets

T	Set of hours of the representative days: the hours of representative day c are the elements from $24 \cdot (c-1)+1$ to $24 \cdot c$, where $c = 1, 2, \dots, C$ and C is the number of representative days considered
T^I	Set of the first hours of the representative days: $T^I = \{24 \cdot (c-1)+1, c = 1, 2, \dots, C\}$
T^L	Set of the last hours of the representative days: $T^L = \{24 \cdot c, c = 1, 2, \dots, C\}$
Z	Set of zones
L	Set of transmission lines, with line $l \in L$ defined by the ordered pair of nodes (z_l^{tail}, z_l^{head}), i.e. a positive flow is from z_l^{tail} to z_l^{head} and a negative flow is from z_l^{head} to z_l^{tail}
FS_z	Set of transmission lines leaving zone z (forward stars)
BS_z	Set of transmission lines entering zone z (backward stars)
K_z	Set of clusters (types) of thermal power plants located in zone z
K	$\bigcup_{z \in Z} K_z$
I	Set of indexes representing distinct offer prices
Q	Set of indexes representing distinct thermal offer prices
K_q	Set of clusters of thermal power plants that offer at the price indexed by q

Parameters

RES:

I_z^V	[€/MW]	Investment cost of solar power plants in zone z
I_z^{WI}	[€/MW]	Investment cost of inshore (onshore) wind power plants in zone z
I_z^{WO}	[€/MW]	Investment cost of offshore wind power plants in zone z
$\overline{Q}^V$	$[MW]$	Standard capacity (size) of solar power plants
$\overline{Q}^{WI}$	$[MW]$	Standard capacity of inshore wind power plants
$\overline{Q}^{WO}$	$[MW]$	Standard capacity of offshore wind power plants

$n^{V}_{z,0}$	$[-]$	Number of solar power plants existing in zone z at the beginning of the planning period
$n^{WI}_{z,0}$	$[-]$	Number of inshore wind power plants existing in zone z at the beginning of the planning period
$n^{WO}_{z,0}$	$[-]$	Number of offshore wind power plants existing in zone z at the beginning of the planning period
$\underline{V}_z$	$[MW]$	Minimum solar power capacity in zone z in the target year
$\overline{V}_z$	$[MW]$	Maximum solar power capacity in zone z in the target year
$\underline{W}_z$	$[MW]$	Minimum wind power capacity in zone z in the target year
$\overline{W}_z$	$[MW]$	Maximum wind power capacity in zone z in the target year
P^{V}	$[€/MWh]$	Offer price of solar power production
P^{WI}	$[€/MWh]$	Offer price of inshore wind power production
P^{WO}	$[€/MWh]$	Offer price of offshore wind power production
$cf^{V}_{z,t}$	$[MWh/MW]$	Solar power capacity factor for zone z in hour t
$cf^{WI}_{z,t}$	$[MWh/MW]$	Inshore wind power capacity factor for zone z in hour t
$cf^{WO}_{z,t}$	$[MWh/MW]$	Offshore wind power capacity factor for zone z in hour t

THERMAL GENERATORS:

I^{K}_{k}	[€/MW]	Investment cost of thermal power plants of cluster k
$\underline{Q}^{K}_{k}$	$[MW]$	Minimum power output of thermal power plants of cluster k
$\overline{Q}^{K}_{k}$	$[MW]$	Maximum power output of thermal power plants of cluster k
$n^{K}_{k,0}$	$[-]$	Number of thermal power plants of cluster k existing at the beginning of the planning period
$\overline{n}^{K}_{k}$	$[-]$	Maximum number of thermal power plants of cluster k in the target year
C^{SU}_{k}	[€]	Startup cost of thermal power plants of cluster k
C^{M}_{k}	$[€/MWh]$	Marginal production cost of thermal power plants of cluster k

P_k^K	[€/MWh]	Offer price of thermal power plants of cluster k
$\gamma_{k,t}^I$	$[-]$	Number of on-line thermal power plants of cluster k at the beginning of hour $t \in T^I$

TRANSMISSION LINES:

$\underline{F}_l$	$[MW]$	Minimum capacity of transmission line l
$\overline{F}_l$	$[MW]$	Maximum capacity of transmission line l

BATTERIES:

I_z^B	[€/MW]	Investment cost of batteries in zone z
$\overline{Q}^B$	$[MW]$	Standard capacity of batteries
$n_{z,0}^B$	$[-]$	Number of existing batteries in zone z at the beginning of the planning period
$\underline{B}_z$	$[MW]$	Minimum battery power capacity in zone z in the target year
$\overline{B}_z$	$[MW]$	Maximum battery power capacity in zone z in the target year
$B_{z,t}^I$	$[MWh]$	Energy content of batteries in zone z at the beginning of hour $t \in T^I$
κ	$[-]$	Loss coefficient for energy stored by batteries ($0 \leq \kappa \leq 1$)
κ^C	$[-]$	Loss coefficient for battery charge ($0 \leq \kappa^C \leq 1$)
κ^D	$[-]$	Loss coefficient for battery discharge ($0 \leq \kappa^D \leq 1$)
rt^C	$[-]$	Rate on battery charge ($0 \leq rt^C \leq 1$)
rt^D	$[-]$	Rate on battery discharge ($0 \leq rt^D \leq 1$)
df	$[-]$	Derating factor ($0 \leq df \leq 1$)

OTHERS:

$D_{z,t}$	$[MWh]$	Load in zone z in hour t

$R_{z,t}$	$[MWh]$	Reserve in zone z in hour t
$\hat{\lambda}_i$	$[€/MWh]$	Offer price $i \in I$
$\hat{\lambda}_q$	$[€/MWh]$	Offer price $q \in Q$ of thermal power plants
M	$[-]$	Total Cost multiplier in the RA constraints ($M = 1$ for cost recovery, $M > 1$ for return on investment)

Variables

UPPER LEVEL

Investment variables:

n_z^V	$[-]$	Number of new solar power plants built in zone z in the target year
n_z^{WI}	$[-]$	Number of new inshore wind power plants in zone z in the target year
n_z^{WO}	$[-]$	Number of new offshore wind generators in zone z in the target year
n_k^K	$[-]$	Number of new thermal power plants of cluster k in the target year
n_z^B	$[-]$	Number of new batteries in zone z in the target year

UC variables:

$\gamma_{k,t}$	$[-]$	Number of thermal power plants of cluster k on-line in hour t
$\alpha_{k,t}$	$[-]$	Number of thermal power plants of cluster k started up in hour t
$\beta_{k,t}$	$[-]$	Number of thermal power plants of cluster k shut down in hour t

Side-payments variables:

u_k^K	$[€]$	Side-payment for thermal power plants of cluster k
u_z^V	$[€]$	Side-payment for solar power plants in zone z
u_z^{WI}	$[€]$	Side-payment of inshore wind power plants in zone z

u_z^{WO}	[€]	Side-payment of offshore wind power plants in zone z

Battery variables

$B_{z,t}$	$[MWh]$	Energy level of battery in zone z in hour t
$B_{z,t}^C$	$[MWh]$	Battery charge in zone z in hour t
$B_{z,t}^D$	$[MWh]$	Battery discharge in zone z in hour t

Variables for price indeterminacy cases

$y_{z,t,i}$	$[-]$	Binary variable equal to 1 if the clearing price in the zone z in hour t is $\hat{\lambda}_i$, 0 otherwise

Variables for avoiding withholding

$\gamma_{k,t}^{SL}$	$[-]$	Number of available thermal power plants of cluster k off-line in hour t
$x_{k,t}^{TF}$	$[-]$	Binary variable equal to 1 if thermal power plants of cluster k do not physically withhold capacity, 0 otherwise

Variables for reserve

$RNP_{z,t}$	$[MWh]$	Reserve Not Provided in zone t in hour t

LOWER LEVEL

$Q_{z,t}^V$	$[MWh]$	Accepted quantity of solar production in zone z in hour t
$Q_{z,t}^{WI}$	$[MWh]$	Accepted quantity of inshore wind production in zone z in hour t
$Q_{z,t}^{WO}$	$[MWh]$	Accepted quantity of offshore wind production in zone z in hour t
$Q_{k,t}^K$	$[MWh]$	Accepted quantity of thermal power plants of cluster k in hour t
$F_{l,t}$	$[MWh]$	Flow on transmission line l in hour t

Dual variables

$\lambda_{z,t}$	$[€/MWh]$	Electricity price in zone z in hour t

$\mu^{V}_{z,t}$	[€/MWh]	Dual variable associated with the total production constraint of solar power plants in zone z in hour t
$\mu^{WI}_{z,t}$	[€/MWh]	Dual variable associated with the total production constraint of inshore wind power plants in zone z in hour t
$\mu^{WO}_{z,t}$	[€/MWh]	Dual variable associated with the total production constraint of offshore wind power plants in zone z in hour t
$\underline{\mu}^{K}_{k,t}$	[€/MWh]	Dual variable associated with the minimum power output constraint of thermal power plants of cluster k in hour t
$\overline{\mu}^{K}_{k,t}$	[€/MWh]	Dual variable associated with the maximum power output constraint of thermal power plants of cluster k in hour t
$\underline{\upsilon}_{l,t}$	[€/MWh]	Dual variable associated with the minimum flow constraint on line l in hour t
$\overline{\upsilon}_{l,t}$	[€/MWh]	Dual variable associated with the maximum flow constraint on line l in hour t

AUXILIARY VARIABLES

$w_{z,t,i}$	$[-]$	Binary variable introduced in the extended formulation
$z^{K}_{k,t,i}$	$[MWh]$	$Q^{K}_{k,t} \cdot y_{z,t,i}$
$z^{V}_{z,t,i}$	$[MWh]$	$Q^{V}_{z,t} \cdot y_{z,t,i}$
$z^{WI}_{z,t,i}$	$[MWh]$	$Q^{WI}_{z,t} \cdot y_{z,t,i}$
$z^{WO}_{z,t,i}$	$[MWh]$	$Q^{WO}_{z,t} \cdot y_{z,t,i}$
$z^{BD}_{z,t,i}$	$[MWh]$	$B^{D}_{z,t} \cdot y_{z,t,i}$
$z^{BC}_{z,t,i}$	$[MWh]$	$B^{C}_{z,t} \cdot y_{z,t,i}$
$n^{K-BIN}_{k,e}$	$[-]$	Binary variable introduced in the mixed-binary version for replacing n^{K}_{k}
$\gamma^{BIN}_{k,t,e}$	$[-]$	Binary variable introduced in the mixed-binary version for replacing $\gamma_{k,t}$
$\alpha^{BIN}_{k,t,e}$	$[-]$	Binary variable introduced in the mixed-binary version for replacing $\alpha_{k,t}$

$\beta_{k,t,e}^{BIN}$	$[-]$	Binary variable introduced in the mixed-binary version for replacing $\beta_{k,t}$
$n_{z,e}^{V-BIN}$	$[-]$	Binary variable introduced in the mixed-binary version for replacing n_z^V
$n_{z,e}^{WI-BIN}$	$[-]$	Binary variable introduced in the mixed-binary version for replacing n_z^{WI}
$n_{z,e}^{WO-BIN}$	$[-]$	Binary variable introduced in the mixed-binary version for replacing n_z^{WO}
$n_{z,e}^{B-BIN}$	$[-]$	Binary variable introduced in the mixed-binary version for replacing n_z^B
s_b^{coef}	$[-]$	SOS2 variable introduced in the piecewise linearization
d_b^{coef}	$[-]$	SOS2 variable introduced in the piecewise linearization
$\overline{s^2}$	$[-]$	Piecewise-linear approximation
$\overline{d^2}$	$[-]$	Piecewise-linear approximation

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

The evolution of electricity systems in recent years all over the world, and notably within the European Community, involves an unprecedented series of transformative changes aimed at reducing carbon emissions and mitigating the impact of climate change, at least according to the interpretation of the anthropogenic nature of the phenomenon. There are several key elements in the evolution of electricity systems concerning decarbonization:

1. **Transition to Renewable Energy Sources**:

    - **Solar and Wind Power**: The significant increase in the deployment of solar and wind power technologies has been a cornerstone of decarbonization. Advances in technology and cost reductions have made these sources increasingly competitive and scalable.
    - **Hydropower and Geothermal**: Traditional renewable sources, including hydropower and geothermal, continue to play a role in the decarbonization process.

2. **Grid Modernization and Smart Technologies**:

    - **Smart Grids**: The evolution of electricity systems involves the integration of smart grid technologies. These systems use advanced sensors, communication networks, and analytics to optimize the performance of the grid, enhance reliability, and accommodate the variability of renewable energy sources.
    - **Demand Response**: Implementing demand response programs within smart grids allows for dynamic adjustments to electricity consumption based on real-time conditions, helping to balance supply and demand.

3. **Energy Storage Integration**:

    - **Batteries and Storage Solutions**: The development and integration of energy storage technologies, such as batteries, are crucial for addressing the intermittent nature of renewable energy sources. Energy storage allows excess energy to be stored and released when demand is high or renewable generation is low, improving grid reliability.

4. **Policy and Regulatory Support**:

    - **Renewable Energy Standards**: Many regions have established renewable energy standards or targets, requiring a certain percentage of electricity to come from renewable sources. These policies provide a regulatory framework to encourage the adoption of clean energy.
    - **Carbon Pricing**: The implementation of carbon pricing mechanisms, such as carbon taxes or cap-and-trade systems, provides an incentive for the reduction of carbon emissions from the power sector.

5. **Phasing Out Coal and Natural Gas**:

    - **Coal Phase-Out**: Governments and utilities are increasingly committing to phasing out coal-fired power plants due to their high carbon intensity. This involves the retirement of existing coal plants and a shift toward cleaner energy sources.
    - **Transition from Natural Gas**: While natural gas has been considered a bridge fuel, efforts are underway to transition away from it in the long term to further reduce greenhouse gas emissions.

6. **Electrification of Other Sectors**:

    - **Transportation**: The electrification of transportation, particularly the widespread adoption of electric vehicles, may contribute to reducing the overall carbon footprint by shifting from fossil fuel-based vehicles to electric power.
    - **Industrial Processes**: Electrifying industrial processes, where feasible, may help reduce emissions from sectors that traditionally rely heavily on fossil fuels.

7. **Decentralization and Distributed Generation**:

- The growth of distributed generation, including rooftop solar panels and small-scale wind turbines, contributes to a more decentralized and resilient electricity system. Localized generation reduces transmission losses and enhances grid flexibility.

8. **International Collaboration**:

    - Global efforts, such as international agreements like the Paris Agreement, underscore the importance of collaborative actions to achieve common decarbonization goals. Countries are working together to share best practices and support each other in transitioning to cleaner energy systems.

9. **Innovation in Advanced Technologies**:

    - Ongoing research and development in advanced technologies, such as next-generation solar and wind technologies, as well as breakthroughs in energy storage and grid management, are critical for achieving deeper decarbonization.

The evolution of electricity systems, with a focus on decarbonization, is therefore an ongoing process of an outstanding complexity, driven by a combination of technological innovation, policy and regulation measures, market dynamics, and international cooperation. The ultimate goal is to create a sustainable, low-carbon energy system that can meet the growing global demand for electricity while minimizing environmental impact.

In this context of evolution, mathematical models play a crucial role in the evolution of electricity systems by providing tools for analysis, planning, optimization, and decision-making. These models help stakeholders understand complex interactions within the electricity system, anticipate future challenges and design effective strategies for sustainable development.

1.1 Thesis motivation and objectives

Generation Expansion Planning (GEP) problems aim at determining future generation plan portfolios in a given Country that are optimal for the operation of the electric energy system in the long term [Conejo et al., 2016]. They generally take the perspective of a fictitious entity, usually called *Central Planner* (CP), that determines capacities, technology, and locations for generators. In recent years GEP problems took the perspective of solving the so-called Trilemma:

- Meet electricity load in the long-term (Energy security)

- Achieve predefined policy targets, such as decarbonization, integration of large shares of renewables, and reduction of CO2 emissions (Environmental sustainability)
- Minimize the sum of the investment and operational cost over the planning horizon (Energy equity)

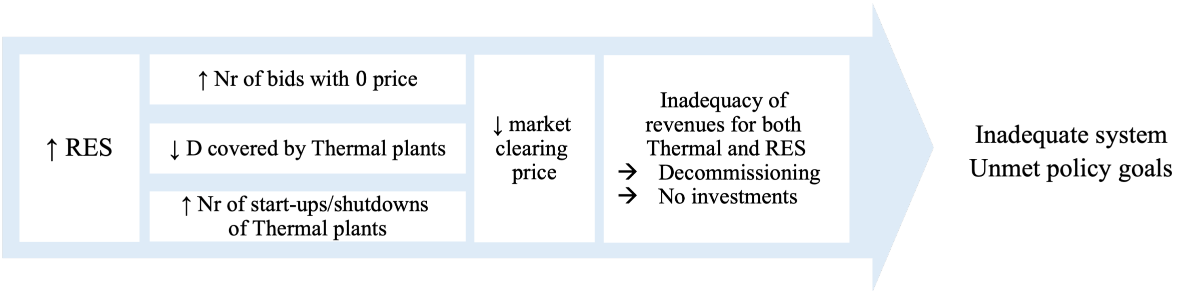


Figure 1.1: Risky evolution of a system with high RES concentration

In the presence of high levels of renewable power sources, the scheduling of programmable power plants (e.g. gas-fired ones) can be strongly influenced by the non-programmable renewable generation as, e.g., the number of startups and shutdowns of the thermal plants can increase, implying a consequential increase in the costs, including O&M costs. In fact, cycling the thermal plants accelerates component wear and tear, resulting in an increase in failure rates, longer maintenance and inspection periods, and higher consumption of spares and replacement components [Rodilla et al., 2013]. To accurately estimate the operative costs, the GEP should consider a good-to-high level of technical details in representing thermal, hydropower and possibly pumping storage, plants, and various kinds of other storage facilities, evaluating power system operation with an hourly resolution which is the typical relevant period length considered in several electricity markets.

The centralized approach provides a benchmark to guide the time-spatial decisions in building new electricity production facilities. Indeed, the actual building of the facilities is carried out by private investors who must be encouraged by policymakers and regulators to follow the social welfare maximizing solution. In providing this benchmark, the quantity-based model proposed by [Micheli and Vespucci, 2020] ignores the *microeconomic* point of view of the individual producer who needs to recover investment and operational costs. This could compromise incentives for new entry of supply resources as well as the incentives of existing units to refrain from retirement. As described in [Hytowitz et al., 2020], future markets with growing penetration levels of renewable energy could have many low-to zero-priced periods, since these power plants exhibit zero marginal costs. In turn, this could affect energy revenues of the resources that are needed for the *long-term system reliability* (i.e. the capacity of the system to meet the load and the reserve margins), preventing them from recovering both operating and capital costs, see Fig. 1.1.

This thesis aims to enhance quantity-based optimization models to provide a benchmark that is *revenue-adequate*. The system configuration will be able to account for energy prices and thus for the economic sustainability of the chosen generators. The prices will be determined endogenously while maintaining their explainability, thus ensuring that investment decisions are truly revenue-adequate. The proposed models will provide guidance on several key aspects, such as investment decisions based on location, technology selection, market clearing prices, and possible payments outside of the market.

Although the centralized approach does not reflect the real dynamics of modern power systems, consisting of several decision-makers involved, the central view is still appropriate to define the target towards which to encourage the system with appropriate interventions. Indeed, even if markets are designed to send signals to critical resources needed for long-term reliability, a mathematical model representing the plurality of independent agents would be highly complex. In fact, it would have to identify among the investment alternatives, those that maximize the profits of each agent over a long-term time horizon, as well as to ensure that the overall demand would be met by the existing generation units. The high dimension of the resulting model would make it not computationally possible to focus on an hourly discretization, which is nevertheless necessary to represent the market mechanism determining the marginal producers and thus the electricity prices in a non-discriminatory market model such as marginal price ones. The idea of having a CP for the auctions for the development of renewables sources has been in fact envisioned by a recent consultation of the Italian Ministry of Environment and Energy Security (MASE), [Ministero dell'Ambiente e della Sicurezza Energetica, 2023]. We shall see later that the proposed model also takes into account possible long-term auctions as those proposed in the aforementioned consultation.

As discussed in [Guo et al., 2022], even though alternatives to centralized coordination have been proposed in the literature in the form of equilibrium problems to account for the strategic actions of the players, it is difficult to conjecture the future degree of market power for each market participant, as long as to set the parameter needed to calibrate those models, and the multiplicity assumptions could strongly impact the results.

Along with the importance for the energy sector, the problem has great relevance in the operations research field, being generally addressed through the formulation of large-scale mixed-integer non-convex models to account for physically-based binary decisions, e.g., lumpy investments, minimum power outputs of production, minimum up and down times, ramps constraints, and startup decisions, e.g. [Van Ackooij et al., 2018] and references therein. Moreover, the price, and revenue, modeling may yield non-convex models requiring dedicated solving algorithms and linearization techniques.

1.2 Introduction to Electricity Markets

Electricity and ancillary services markets have several peculiarities around the world and notably within the European Community. Fig. 1.2 depicts a generic scheme of electricity and balancing markets, including bilateral contracts and several interactions among different entities, such as Market Operator (MO), Transmission System Operator (TSO), producers owing programmable and non-programmable production units, loads, and traders-wholesales.

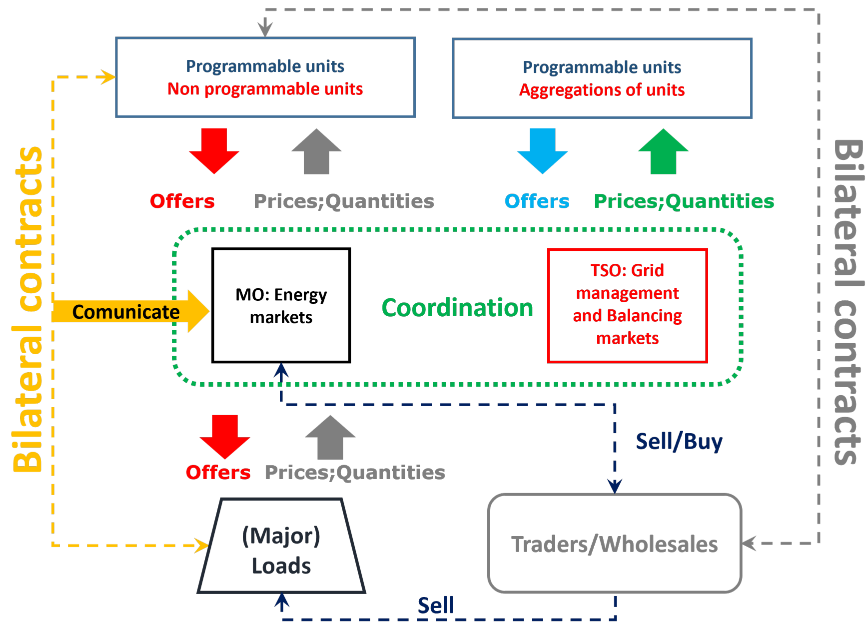


Figure 1.2: Generic electrical energy markets scheme

With the generic scheme of Fig. 1.2 in mind, we briefly describe the specific market sequence in Italy where the day-ahead market (MGP) is included in the European coupling design as *per* EC Regulation 1222/2015, *Capacity Allocation and Congestion Management* (CACM), while the ancillary services market by Regulation 943/2019 and balancing Regulation. In Fig. 1.3 we sketch the Italian cascading markets, from forward long-term (MTE) to balancing (MB) real-time markets, up to the European balancing platforms (PBil) such as MARI and PICASSO. In the present thesis, we will be focusing

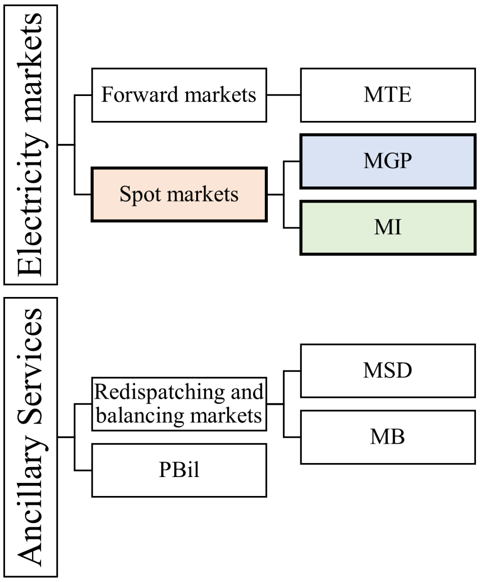


Figure 1.3: Italian Electrical Energy Markets

on the spot electricity markets.

1.3 The missing money problem

Within liberalized electricity markets, the *missing money problem* materializes in the long term when generators fail to recover their investment (i.e. CAPEX) and operational costs through market revenues coming from electricity, and possibly ancillary, services. There are two main causes of the missing money problem:

1. insufficient scarcity rents
2. non-convexities

The first case (1) is possible for both convex and non-convex models. In a pay-as-cleared

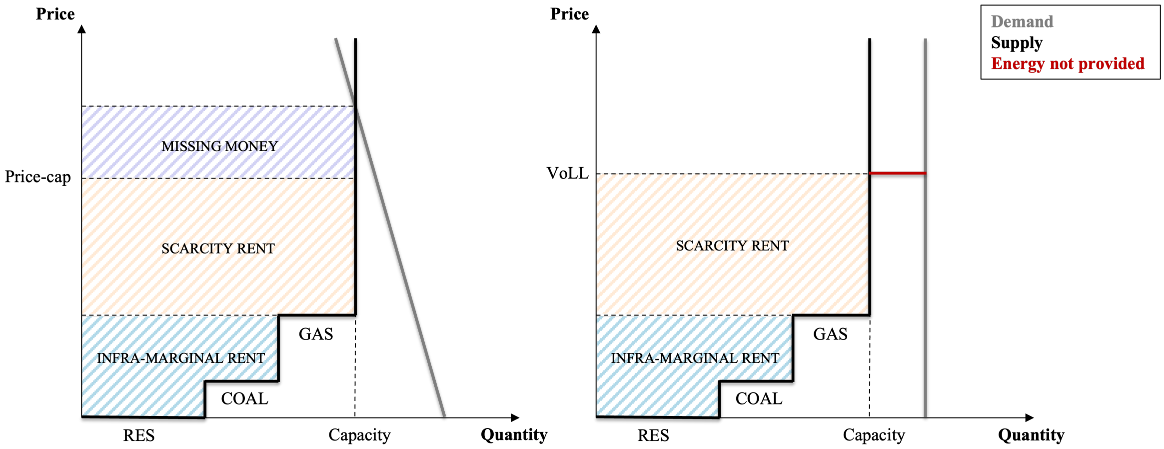


Figure 1.4: Infra-marginal and Scarcity rents with elastic (left) and inelastic (right) demand

mechanism[1], the market clearing price is set by the intersection of the demand and supply curves, the latter constructed according to the merit order. This price is assigned to each accepted generator. The marginal generator is the one who sets the price, while the infra-marginal generators, having bid at lower prices, receive an *infra-marginal* rent. As illustrated in Fig 1.4, when the offered capacity is not sufficient to cover the demand, market prices spike up and are determined either by the - elastic - demand itself or by the *Value of Lost Load* (VoLL), however defined within the specific market in the case of - inelastic - demand [Boiteux, 1960]. The portion of the revenue attributable to the difference between the market price and the cost of the marginal generator is called *scarcity rent* and should be considered as revenue to cover the investment costs of both infra-marginal and marginal generators [Bowring and Tyler, 2019].

Insufficiency of scarcity rent could occur either when the price is administratively capped to a low value due to market concerns [Byers and Hug, 2023] or when the number of times in which the price reaches the VoLL is zero or not enough. This leads to a shortfall in revenue for existing generators and an insufficient incentive to invest in new ones. This is the case of the *Energy-only market* (or *Reserve and Energy market*) which does not compensate for capacity and therefore carries the risk that generation companies will delay investment because they do not know whether and to what extent the price of electricity will be set at a level sufficient to cover their costs [Fraunholz et al., 2023]. *Capacity Remuneration Mechanisms* (CRM) can be introduced to address the missing money problem related to the insufficiency of scarcity rents. However, a criticism in [Mays et al., 2021] relates to the potential misallocation of resources and distortion of the capacity mix, as these mechanisms are generally not capable of capturing real-time information about operating conditions. CRM can be classified into different categories such as those depicted in Fig. 1.5.

Notably, to ensure the **adequacy of the system**, with Decision n. ARG/elt/98/11 of the Italian Authority Regulatory for Energy, Networks and Environment (ARERA), Italy has implemented a *Capacity Market* for thermal generators, where the TSO (Terna) defines competitive auctions, in which:

- the demand curve represents the system's willingness to pay for capacity according to the adequacy targets set for the Italian electricity system, three hours
- the supply curve is defined by participants submitting their bids, which are defined

[1] As opposite to the pay-as-bid mechanism, where the price captured by each accepted generator equals its bid

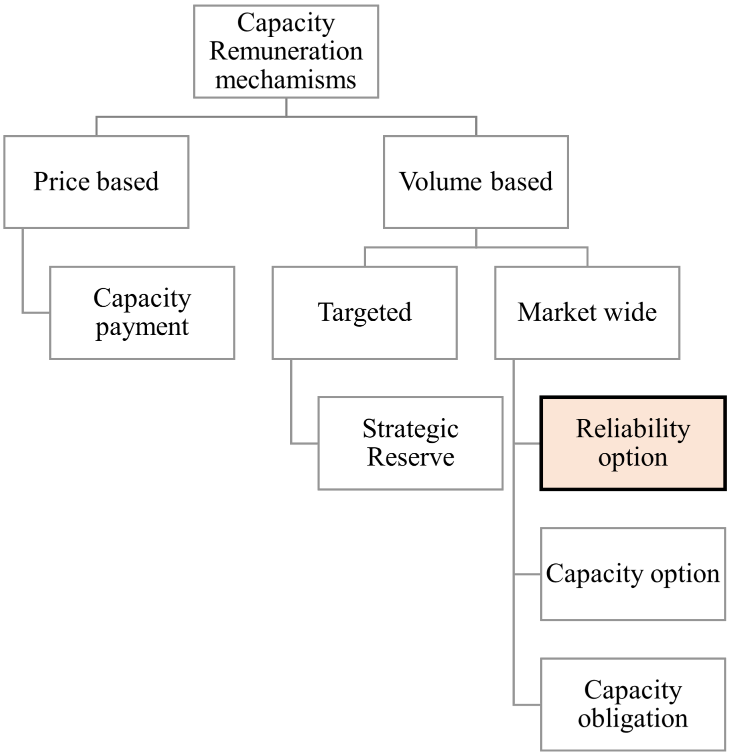


Figure 1.5: Capacity Remuneration Mechanism Taxonomy

by a premium in power [€/MW/year] for a certain time horizon and a capacity [MW/year]

Once the auction clears a unique premium in [€/MW/year] is defined for each accepted bid based on the intersection of demand and supply curves. Generators providing the allocated capacity receive the fixed annual premium from Terna and pay back the difference, if positive, between the price realized on the energy market (MGP), and ancillary services markets (MSD/MB), and the *strike price* defined by the ARERA based on the marginal cost of an efficient open cycle gas turbine (OCGT), Decision n. 399/2021/R/eel and subsequent. Thus realizing mandatory one-way[2] Contracts for Difference (CfD). The Italian design for the Capacity Market is typically referred to as a *reliability option*-type of CRM. For an overview of the EU CRM and the theoretical underpinning of CRM, we refer to [Papavasiliou, 2021]. Renewable generators, instead, to stabilize revenue streams and thus increase the bankability of the projects can currently rely on two-ways CfD: for RES the CRM-like auction can be interpreted as an instrument to achieve the decarbonization targets while ensuring the financial stability of the developments.

Additionally in the recent consultation on the Electricity Market Design, the European

[2]one-way CfD envision a single strike price or cap, while two-ways CfD envision a cap and a floor, therefore in this latter case the producer is hedged against lower, below floor, marginal prices

Commission envisions two-ways CfD as the main instrument to promote investments so as to reach decarbonization. However, classic and simple two-ways CfD are criticized by the FERX consultation [Ministero dell'Ambiente e della Sicurezza Energetica, 2023] as interfering with market price signals:

- the producer ignores zonal construction criteria and hourly bidding strategies generally used to obtain a higher day-ahead price: since the producer receives the auction price for each unit produced regardless of the zone and the time of the bid, it will only act with the aim of maximizing its production.

- the producer bears the revenue risk related to the volume. When selling electricity at market prices, the negative correlation between prices and the availability of wind/solar power helps to mitigate the revenue risk (lower wind/solar production corresponds to higher prices). When the revenues are based on CfD, on the other hand, the reduction in production volume does not affect its captured price. Additionally, there is a volume risk associated with production reductions imposed by overgeneration or curtailment.

Therefore, the consultation proposes two potential evolutions of CfD also towards the so-called smart CfD that somehow face the dispatch efficiency issue. Additionally, the consultation proposes two main approaches to the auctions:

a. centralized asset-based model

b. decentralized model with standard profiles

These alternatives are put forth with the objective of attaining decarbonization targets while ensuring cost-minimization for consumers, stimulating efficient resource investment and operation, and enhancing risk distribution among the involved players.

The centralized model (a) considers a centralized optimization process where the system, the central planner, defines quantity, location, and types (the expected production profile) of renewable sources to be built. In the decentralized model (b), instead, the system defines only quantity and location, leaving the choice of the technologies' portfolio to the producers. At the same time, the system requires certain standard production profiles from the producers.

The second case (2) refers to the presence of non-convex costs (e.g., startup costs[3])

[3]that account for the fuel needed to raise the boiler to its minimum operating temperature prior to producing electricity

and technical constraints (e.g., minimum generation requirements and min up and down time constraints). These non-convexities make it difficult, if not impossible, to determine a spot price that precisely captures all the costs, that additionally are non-separable along the time, e.g. [Taylor, 2018].

To cope with these inherent difficulties, several approaches have been proposed, one of the most promising introduced the concept of Convex Hull Price (CHP), which was first suggested in [Gribik et al., 2007] — and its approximations, often called *Extended Locational Marginal Prices* (ELMPs) [Wang et al., 2016], see also [Stevens and Papavasiliou, 2022]. In essence, the CHP approaches solve a relaxation as close as possible to the non-convex formulation (the Convex Hull) and deduce prices from it. Lagrangian relaxations or specialized formulations of constraint blocks that inherently have the property of being a convex hull are often used, for example on the minimum up and down times, [Rajan et al., 2005]. In fact, as explained in [Ruiz et al., 2012], binary variables are needed to represent physically-based non-convexities but their presence prevents obtaining marginal prices as dual variables of balance equations. This makes, in turn, difficult to define an adequate remuneration for the units, leading the literature to propose several pricing schemes. Since the difference between remunerations, as defined by the specific pricing scheme, and operation costs finances the capital costs of infra-marginal units, different investment decisions should be expected under each pricing scheme [Herrero et al., 2015].

In Section (2.1) we will detail several alternative pricing formulations proposed by the literature also to deal with the missing money problem in the short term.

Chapter 2

Literature review

This chapter presents an extensive literature review of the most recent contributions in addressing the revenue adequacy issue for generators, highlighting the main open problems. In Section (2.1), we explore clearing models in electricity markets, paired with various pricing models, a fundamental step for both grasping the studies presented in Section (2.2) and for appreciating our modeling choices described in the subsequent chapters. Indeed, the papers reviewed consider different pricing schemes, the understanding of which is essential for identifying the gaps this thesis aims to address. Among these gaps are the lack of explainability and simplicity of the prices obtained and the impossibility of applying some pricing schemes in optimization problems, as opposed to heuristic approaches which, as such, do not guarantee obtaining the most efficient solution. Moreover, the revenue adequacy is guaranteed for specific technologies only, which may result in significant losses for others or not coherent choices, since for example the investment decisions in renewables and batteries are strictly related. Equally important, the analysis conducted has been instrumental in understanding how to account for startup costs in our modeling (thereby ensuring these costs are also covered) and the impact of non-convexities on the market outcomes. Furthermore, the pricing section paves the way for introducing the reader to strategic mechanisms, some of which will be addressed by our model.

2.1 Pricing schemes literature review

For a meaningful evaluation of revenue-adequate investments, it is essential for the models to consider proper energy pricing [Frew et al., 2016]. The "right" prices provide appropriate economic signals in both the planning and operation phases of real markets [Eldridge et al., 2020], [Mays et al., 2021], in fact:

- in the long term, they efficiently coordinate entry (building) and exit (decommissioning) of different types of generators, thereby promoting an optimal resource mix, i.e., the resource mix that satisfies system constraints at the least cost. This is true at least independently of Policy indications such as decarbonization;

- in the short term, they support the least-cost dispatch, thus guaranteeing market participation.

Moreover, short-term prices should give signals based on the load (higher price with higher demand) to encourage the units with limited capacity, such as storage hydropower plants and batteries, to bid in hours of peak, increasing the efficiency of the market by performing peak-shaving without using other expensive units [Frangioni and Lacalandra, 2022]. Additionally, correct short-term pricing can be useful for demand response's evolution, e.g. [MacDonald et al., 2023]. The work in [Coutu and White, 2014a] illustrates three key principles that a market and its pricing formation process should satisfy:

1. Efficiency
2. Price Transparency
3. Simplicity

The first principle (1) consists of identifying the dispatch that minimizes the production costs corresponding to the offered prices and for which the dispatched generators do not want to deviate, since any deviation would put them in a less favorable position. The second principle (2) relates to a situation where market participants are aware of the prices received by others and also of the replicability of the market outcomes. Finally, the third principle (3) involves minimizing the number of prices, ideally having as few as possible for each location and time, and having a simple logic that buyers/sellers understand. This means avoiding complexities that might pose challenges in interpreting the price. These three principles were a guideline for instance in designing the Italian MGP.

We will begin by illustrating the clearing methods in the US (2.1.1), where the system operator allows participants to explicitly declare non-convexities in their multi-part bids. Subsequently, we will delve into European markets, with a focus on the Italian day-ahead market (2.1.2). While in some EU markets, complex bids (e.g., block orders in the Nordic-Baltic region, minimum income conditions in Spain) are considered to address non-convexities, Italy considers simple bids. In fact, the Italian market operator ignores all the non-convexities on the day-ahead market, leaving the target of obtaining a production profile that is consistent with the technical constraints and economically viable to the bidding skills of each generation company and to the intra-day market, which is specifically designed in order to cope with possible unfeasible programs.

2.1.1 US market clearing

The US configuration with multi-part bids is considered: generators share all technical constraints (e.g. technical minimum production, minimum up and down times) and costs (marginal and startup costs) with the central system operator, who clears the market. In Tab. 2.1 (adapted from [Herrero et al., 2020]) key offer parameters are highlighted.

Operating Costs		Technical Constraints	
Quantity, Marginal Cost	MWh, $/MWh	Minimum, Capacity	MWh
No-load Cost	$	Ramp Rate	MWh/h
Start-up Cost	$	Min Max Up Time	h
		Min Down Time	h

Table 2.1: Typical multi-part offer parameters

The process of clearing the market and evaluating the outcomes typically involves three steps, each corresponding to an optimization model, as depicted in Fig. 2.1:

- Unit Commitment and Economic Dispatch (UCED), which establishes the least-cost dispatch (statuses and productions);
- PRICING, which determines the clearing prices;
- Profit Maximization (PM), which is employed to conduct an analysis of the losses that a generator may perceive by following the least-cost dispatch.

The first modeling step is called *Unit Commitment and Economic Dispatch* (UCED). Given the demand d (inelastic), the costs (C_k^F, C_k^M) and the physical characteristics $(\underline{Q}_k, \overline{Q}_k)$ for each generator k, the **UCED** consists in solving the least-cost problem (2.1):

$$UCED(d) = \min_{Q_k,\gamma_k} \quad \sum_k (C_k^F \gamma_k + C_k^M Q_k) \tag{2.1a}$$

$$\text{s.t.} \quad \sum_k Q_k = d \tag{2.1b}$$

$$\underline{Q}_k \gamma_k \leq Q_k \leq \overline{Q}_k \gamma_k \qquad k \in K \tag{2.1c}$$

$$\gamma_k \in \{0,1\} \qquad k \in K \tag{2.1d}$$

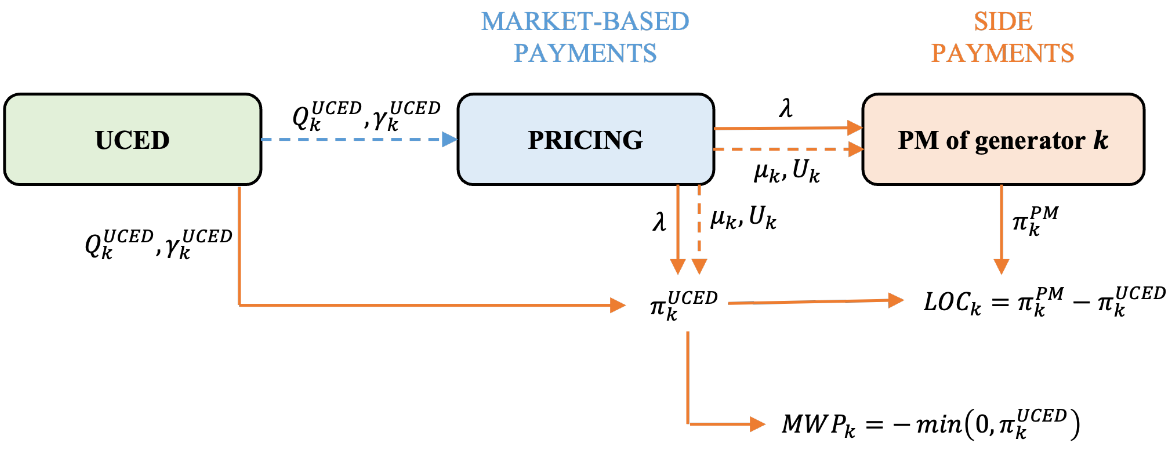


Figure 2.1: Steps for determining payments: the rectangles refer to the optimization models used in the specific step. The continuous lines indicate the values that are always used as inputs, while the dashed lines indicate inputs that are only required by some pricing approaches. Specifically, for each generator k the profit π_k^{UCED} is computed given the accepted quantity Q_k^{UCED} and the status γ_k^{UCED} from UCED and the price λ (and possibly other compensations μ_k, U_k) from PRICING. The make-whole payments MWP_k are calculated as the payments needed to cover losses. Given the price, the maximum profit π_k^{PM} is determined by PM and then the lost opportunity costs LOC_k are calculated.

(2.1) is a mixed-integer problem (MIP) that determines for each generator $k \in K$ the **status** $\gamma_k \in \{0, 1\}$ and the **production** $Q_k \geq 0$ so as to satisfy

- the system-wide constraint: supply-demand balance constraint (2.1b)
- local constraints for each $k \in K$: technical minimum and capacity constraints (2.1c)

and to minimize the system cost (2.1a) to supply demand d.

The function $UCED(d)$, which represents the value of the least cost solution as demand d varies, is called *value function* [Gribik et al., 2007].

Let's consider the example in Tab. 2.2 and a demand $0 < d \leq 40$.

k	C_k^F	C_k^M	$\underline{Q}_k$	$\overline{Q}_k$
	[\$]	[\$/MWh]	[MW]	[MW]
A	70	5	0	15
B	0	10	11	15
W	0	0	0	10

Table 2.2: Economical and physical characteristics of the generators considered in the example

The system configuration has three generators A, B, C with marginal costs of 5, 10, and 0 \$/MWh respectively; generator A has a cost of 70 \$ associated with its startups while generator B presents a technical minimum level of 11 MW. The value function $UCED(d)$,

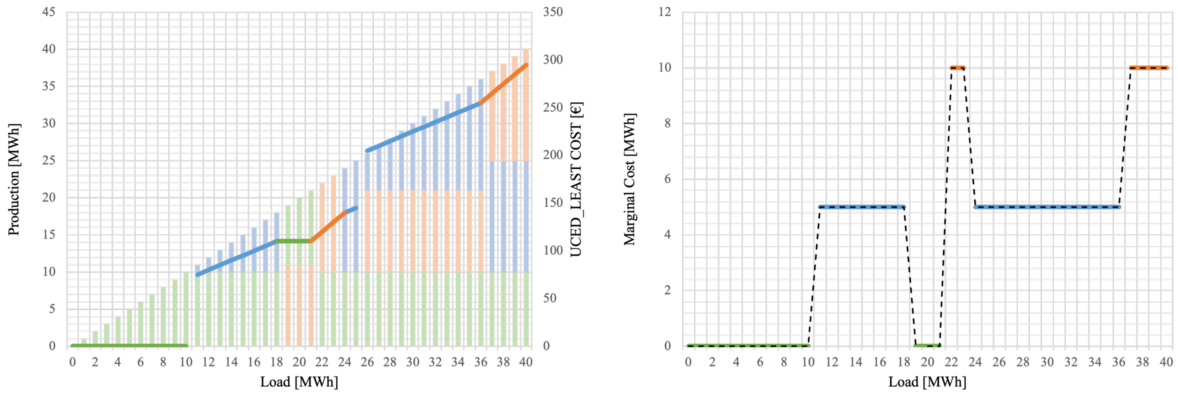


Figure 2.2: Left: Value function $UCED(d)$ (continuous lines) and dispatched quantities $Q_k(d)$ $\forall k$ (columns), for different levels of load d. Right: Slope of $UCED(d)$, representing the marginal cost of the marginal generator. In both the figures, the colors refer to the units (A, B, W).

.

depicted in Fig. 2.2 by continuous lines, is discontinuous in one point and presents a strong non-monotonically increasing slope. The swinging slope represents *the marginal cost of the marginal generator*, i.e., the generator that would cover an additional unit of load. These s are determined by the non-convex feasible sets X_A, X_B of the generator A and B respectively, defined by their local constraints. In particular, the discontinuity is caused by the positive value of the technical minimum $\underline{Q}_B$ of generator B while the slope is affected by the positive fixed cost C_A^F of generator A. Due to the non-convexities, for some levels of demand the marginal generator is not the most expensive committed generator. For example, with a demand $d = 20$, the least-cost solution requires the production of generators W and B for 9 MWh and 11 MWh respectively. The slope of the least-cost function assumes a value of 0, albeit the most-expensive unit committed is B with a marginal cost of 10, as the extra unit of demand would be met at the least cost by W. Moreover, in the discontinuities, the slope is not defined and does not reflect the variation in the cost associated with an additional unit of load (with respect to the least-cost value) as the fixed cost and the discontinuities are not captured. For example, with $d = 9$ the slope is 0 while with $d = 11$, the slope is 5; the value function for each value of the demand is $UCED(9) = 0, UCED(10) = 75, UCED(11) = 80$. $UCED(10) - UCED(9) = 75$ is not reflected by the marginal cost.

The of the slope representing the marginal cost (which increases and then decreases, and then increases again as the load increases) and the resulting operating losses for the generators do not allow the traditional definition of

price equals marginal cost of the marginal generator

to be used and raise the question of how to define the **PRICING** problem that determines the market-based prices, with consequent payments within the market.

A pricing scheme can be classified as

- *uniform* (also called *linear* or *non-discriminatory*), if each dispatched generator receives the same market-based payment;
- *multi-part* (also called *non-linear* or *discriminatory*), if the market-based payment differs between generators.

Discriminatory payments can be used to achieve economic efficiency and to signal the need for economic investments while non-discriminatory payments can result in a too broad distribution of the costs over the uniform price to provide a meaningful signal [O'Neill et al., 2017].

The pricing scheme can be also referred to as *volatile* if a small change in load leads to a large change in prices [Bjørndal and Jörnsten, 2008].

We will shortly review the different PRICING schemes that are currently used in the US.

A *competitive partial*[1] *equilibrium* (or *Walrasian equilibrium*) is defined by a set of market-based prices (which we will refer to as λ, and eventually as μ_k, ρ_k if additional discriminatory market-based payments are envisaged) and least-cost dispatch ($Q_k^{UCED}, \gamma_k^{UCED}$ for each generator k) such that, [O'Neill et al., 2005], [Schiro et al., 2016]:

a. the market clears (the demand is satisfied by the dispatched production);

b. every participant maximizes its utility given the payments it receives.

In [Motto and Galiana, 2002] it is considered the additional condition "consumers have no budget constraint" for the existence of a competitive equilibrium.

The condition (a.) is in general fulfilled by $Q_k^{UCED}, \gamma_k^{UCED}$ because of the balance constraint (2.1b); however, a competitive market clearing price does not exist when there is no intersection of demand and supply curves: see the example in [Araoz and Jörnsten, 2011].

[1] Only one market is considered.

The condition (b.) is evaluated by considering a **Profit Maximization** (PM_k) problem for each generator k as in (2.2).

$$\pi_k^{PM} = \max_{Q_k,\gamma_k} \quad \lambda Q_k - C_k^F \gamma_k - C_k^M Q_k \tag{2.2a}$$

$$\text{s.t.} \quad \underline{Q}_k \gamma_k \leq Q_k \leq \overline{Q}_k \gamma_k \tag{2.2b}$$

$$\gamma_k \in \{0,1\} \tag{2.2c}$$

Given the price λ (obtained from the chosen PRICING problem), with (2.2) the generator k determines the maximum profit it could obtain if it were able to self-schedule, taking into account its internal constraints (2.2b); note that the level of load d is not taken into account in this problem.

Then the difference between the maximum possible bid-in profit π_k^{PM} and the profit obtained when following the least-cost dispatch

$$\pi_k^{UCED} = \lambda Q_k^{UCED} - C_k^F \gamma_k^{UCED} - C_k^M Q_k^{UCED}$$

can be calculated so as to check condition (b.): if $\pi_k^{PM} = \pi_k^{UCED}$ for each generator k, (b.) is satisfied and the prices support the least-cost solution. In particular, the market-based payments are sufficient and *market-based incentive compatibility* exists; this condition is called *profit optimality*.

However, *profit-suboptimality* may exist when a generator prefers a production schedule that differs from the least-cost dispatch, as it can obtain a greater profit $\pi_k^{PM} > \pi_k^{UCED}$. The generator bears losses, known as *Lost Opportunity Costs* (LOC), equal to the difference between its preferred profit and what it would earn at the same price following the least-cost solution:

$$LOC_k = \pi_k^{PM} - \pi_k^{UCED}$$

Generators who bear LOC can learn to increase profits e.g. through a reinforcement learning algorithm: they adopt a strategic bidding behavior consisting, for example, of bidding with zero fixed operating costs (self-commitment) or zero total costs (self-scheduling) to maximize the quantity accepted, exercising market power by becoming price takers [Byers and Eldridge, 2022].

Side payments equal to the LOC are needed so that generators have no incentive to deviate from the UCED solution.

The system operator can provide side payments that only cover a part of the LOC: *Make-Whole Payments* (MWP) ensure that the generator k receives at least its cleared bid-in costs [Schiro et al., 2016]:

$$MWP_k = -\min\{0, \pi_k^{UCED}\}$$

Since the inclusion of a unit in the optimal dispatch implies that its presence improves the objective function value, there always exists enough surplus in the market to support the MWP [Mays et al., 2021]. The latter is not true for LOC: see the example in [Madani and Papavasiliou, 2022].

Considering transmission and reserve constraints, additional side payments, called *product revenue shortfalls* [Eldridge et al., 2020] (or excess product payments or financial transmission rights uplifts [Schiro et al., 2016]), could be required to prevent underfunding of ancillary service providers and financial transmission rights holders; those payments are needed when different binding inequality constraints exist in the UCED and in the PRICING problems.

Incentive compatibility exists if the incentive to deviate from the least-cost dispatch is eliminated by the combination of market-based and side payments.

Note that side payments should be reduced as they are discriminatory and not transparent, making generators not able to be hedged in the forward markets [Schiro et al., 2016] and distorting market entry incentives [Eldridge et al., 2020].

Mathematically, the previous concepts can be illustrated as follows [Pablo Luna et al., 2021]: when the market-based price λ alone supports the least-cost solution, the following inequality is valid

$$UCED(d') \geq UCED(d) + \lambda(d' - d), \forall d' \tag{2.3}$$

and requires the affine function $UCED(d) + \lambda(d' - d)$ to stay below the value function $UCED(d')$. Therefore,

- if $UCED(d)$ is convex, the set of supporting prices (more than one price in case of price multiplicity, which will be discussed below) for a given demand d coincides with the subdifferential

$$\partial\ UCED(d) = \{\lambda : UCED(d') \geq UCED(d) + \lambda(d' - d),\ \forall d'\}$$

- if $UCED(d)$ is non-convex (as in the previous example), the existence of a price λ that alone supports the least-cost solution is rare.

Additionally, in the points where $UCED(d)$ is discontinuous, the price λ can take any value in a set that is unbounded.

We now focus on the second block of Fig. 2.1, illustrating different PRICING models that have been proposed. It is crucial to emphasize that, even though the dispatch determined by PRICING may differ from that of UCED, only the price information from the PRICING block is considered relevant. This does not apply to the P-D approach (2.26), which determines price and quantity in one step.

2.1.1.1 Restricted

The simplest PRICING approach is the one called *Restricted* that was proposed in [Gribik et al., 2007], also known as *Locational Marginal Price* [Eldridge et al., 2018] or *Fixed Configuration Pricing* [Byers and Hug, 2023].

After solving the UCED problem (2.1) to find the least-cost dispatch ($Q_k^{UCED}, \gamma_k^{UCED}\ \forall k$), the **RESTRICTED MODEL** (R) (2.4) is considered:

$$\min_{Q_k} \quad \sum_k (C_k^F \gamma_k^{UCED} + C_k^M Q_k) \tag{2.4a}$$

$$\text{s.t.} \quad \sum_k Q_k = d \qquad (\lambda^R) \tag{2.4b}$$

$$\underline{Q}_k \gamma_k^{UCED} \leq Q_k \leq \overline{Q}_k \gamma_k^{UCED} \qquad k \in K \tag{2.4c}$$

(2.4) is a linear problem (LP) obtained from the UCED problem by fixing γ_k with optimal statuses γ_k^{UCED}. The optimal value of the dual variable λ^R associated with the supply-demand balance constraint (2.4b) is the **electricity (R) price**. At price λ^R, the MWP can be calculated, together with the LOC after solving the PM problem (2.2).

It is worth noting that R replicates the least-cost dispatch ($Q_k^R = Q_k^{UCED}\ \forall k$) and produces prices λ^R exactly equal to the marginal costs of the marginal generators described above. Therefore, the prices are volatile, as they do not monotonically increase with demand, posing challenges for demand-response program incentives [Pablo Luna et al., 2021] and causing financial outcomes instability [Gribik et al., 2007].

The electricity price λ^R is exclusively determined by the marginal cost of the resource with $\gamma_k^{UCED} = 1$. Certain generators that are set off-line by UCED may perceive LOC as the startup and operation at full power would result in a profit that exceeds the zero profit obtained by following the least-cost dispatch; this occurs when λ^R is sufficient to cover both the marginal and the fixed costs of the generator. A side payment equal to the LOC is then required to make the generator indifferent between starting up and following the least-cost dispatch. Other online generators may instead realize an infra-marginal profit when a more expensive generator sets the price. As the dual variable associated with the balance constraint, the price represents the variation in the objective function associated with an additional unit of load. Therefore, it reflects neither the fixed costs, since the constant $C_k^F \gamma_k^{UCED}$ can be removed from the objective function without changing the optimal solution, nor the marginal costs of the units operating at the technical minimum, since in this case an additional unit of demand can generally be met by a less expensive generator (that still has residual capacity). Side payments equal to the economic losses are required to make the generators whole. At points of discontinuity in the value function $UCED(d)$, λ^R is any value between the highest marginal cost of online generators and the VoLL (price multiplicity). In such instances, the level of load precisely equals the sum of the capacities of online generators. It is noteworthy to mention that off-line generators are effectively treated as if they do not exist, and as such, they cannot define the upper bound of the price interval.

2.1.1.2 Integer programming

Since R requires high side payments to cover economic losses and LOC, the following *Integer Programming* approach has been proposed in [O'Neill et al., 2005]. It is also called *post-UC problem* [O'Neill et al., 2017]. After solving the UCED problem (2.1) to find the least-cost dispatch ($Q_k^{UCED}, \gamma_k^{UCED}\ \forall k$), the **INTEGER PROGRAMMING**

MODEL (IP) (2.5) is considered:

$$\min_{Q_k,\gamma_k} \quad \sum_k (C_k^F \gamma_k + C_k^M Q_k) \tag{2.5a}$$

$$\text{s.t.} \quad \sum_k Q_k = d \qquad \lambda^{IP} \tag{2.5b}$$

$$\underline{Q}_k \gamma_k \leq Q_k \leq \overline{Q}_k \gamma_k \qquad k \in K \tag{2.5c}$$

$$\gamma_k = \gamma_k^{UCED} \qquad k \in K \quad (\mu_k^{IP}) \tag{2.5d}$$

$$\gamma_k \geq 0 \qquad k \in K \tag{2.5e}$$

(2.5) is a LP obtained from the UCED problem by adding constraints (2.5d) [2] that assign the optimal statuses γ_k^{UCED} to status variables γ_k, redefined as real non-negative. The optimal value of the dual variable λ^{IP} associated with the supply-demand balance constraint (2.5b) is the **electricity (IP) price** while the optimal value of the dual variable μ_k^{IP} is the **commitment price** of the generator k; λ^{IP} measures the rate of change of the objective function associated with the balance constraint, keeping the commitment decisions fixed, while μ_k^{IP} measures the rate of change of the objective function associated with the integrality nature of the commitment, and therefore, the impact of moving away from the original 0-1 decision [Pablo Luna et al., 2021]. The revenues for electricity and commitment are given by

$$\pi_k^{UCED} = \lambda^{IP} Q_k^{UCED} + \mu_k^{IP} \gamma_k^{UCED} - (C_k^F \gamma_k^{UCED} + C_k^M Q_k^{UCED})$$

The market-based payment for generator k is expressed as $\lambda^{IP} Q_k^{UCED} + \mu_k^{IP} \gamma_k^{UCED}$. It is important to note that the term $\mu_k^{IP} \gamma_k^{UCED}$ is only defined when $\gamma_k^{UCED} = 1$. It represents a payment received by generator k if $\mu_k^{IP} > 0$ or a payment made by generator k to the auctioneer if $\mu_k^{IP} < 0$. Positive values compensate for economic losses, while negative values nullify infra-marginal profits, thus penalizing efficiency [Van Vyve, 2011] and undermining the financing of capital costs. Consequently, $\pi_k^{UCED} = 0$ for each generator k, and MWP are not required. By nullifying losses, IP can incentivize participants to bid untruthfully, knowing that their stated costs are guaranteed to be reimbursed (as in the pay-as-bid auction) [Byers and Hug, 2023]. Positive values render the generator indifferent to being committed or not, while negative values can function as a penalty to

[2] The sign of these equalities has to be chosen such that the associated dual variables assume the right sign to nullify π_k^{UCED}.

disincentive the startup of the plant. Consequently, the maximum profit $\pi_k^{PM} = 0$ for each generator k and side payments are not necessary to support the least-cost solution. In other words, market-based incentive compatibility exists [3]. The positive payments are usually assumed to be distributed among consumers on a pro-rata basis, treating consumers unfairly compared to generators, while the negative payments are only potential to discourage the generators from deviating from the least-cost solution.

IP replicates the least-cost dispatch ($Q_k^{IP} = Q_k^{UCED}\ \forall k$) and the electricity prices are volatile, being equal to the one determined by R ($\lambda^{IP} = \lambda^R$). Given the existence of distinct commitment prices for different generators, it follows that IP is discriminatory.

To tackle the challenges associated with IP, two adjusted approaches have been introduced. The **Integer Programming Positive** (IP+) approach proposed in [O'Neill et al., 2005] aims to address the issue of nullifying infra-marginal profit. The electricity price remains the dual variable λ^{IP}, while for the commitment price, $\mu_k^{IP+} = \max(0, \mu_k^{IP})$ is considered. As a result, generators receive payments for their commitment to cover economic losses but retain infra-marginal profits. By assuming only non-negative values, commitment prices either make the generator indifferent to being committed when it suffers losses or increase the maximum profit the generator can achieve, thereby creating LOC. The **Modified Integer Programming** (modIP) proposed in [Bjørndal and Jörnsten, 2008] was designed to mitigate both volatility and discriminatory concerns. Viewing the IP problem (2.5) as a Benders sub-problem, it adds extra constraints that fixe some continuous variables, referred to as *price-complicating variables*, at their optimal values, so that prices are piece-wise constant and non-decreasing in d. In addition to the electricity price, generators receive a non-discriminatory fixed payment determined on the basis of the dual variables associated with the additional constraints. For more details, see the electronic companion to [Liberopoulos and Andrianesis, 2016].

The approaches described in the paragraphs below address other problems related to IP or its variants. The pricing schemes CH (2.6),(2.11), have been suggested to address the significant side payments that may be necessary with the IP+ and modIP approaches. The objective of these schemes is to determine prices, among all possible uniform ones, that minimize the side payments needed to support the least-cost solution. The GU approach (2.25) redistributes the cost of reaching profit optimality: both generators and consumers can be charged to provide subsidies to others. The sum of these transfers

[3] Only in the absence of scarcity [Byers and Hug, 2023]; a price equal to the VoLL may create LOC.

is null, ensuring the *system operator's neutrality*, i.e., the condition for which the total payment collected from the consumers is exactly equal to the total remuneration paid to the generators. The P-D approach (2.26) and various market power mitigation measures, including deviation penalties, can be used to solve the problem of bid non-truthfulness.

2.1.1.3 Convex Hull

For small-scale examples (single time interval, single node), the application of a *graphical Convex Hull* method is suitable to help intuition on this approach. Starting with the UCED problem (2.1), this method graphically represents the value function $UCED(d)$, which we call ψ, and its convex hull, conv ψ. In this context, the conv ψ is defined as the point-wise greatest convex function majorized by the value function ψ. It can be graphically obtained by connecting the extreme points of the non-convex steps of $UCED(d)$ with straight lines, as illustrated in Fig. 2.3. The slope of the convex hull of

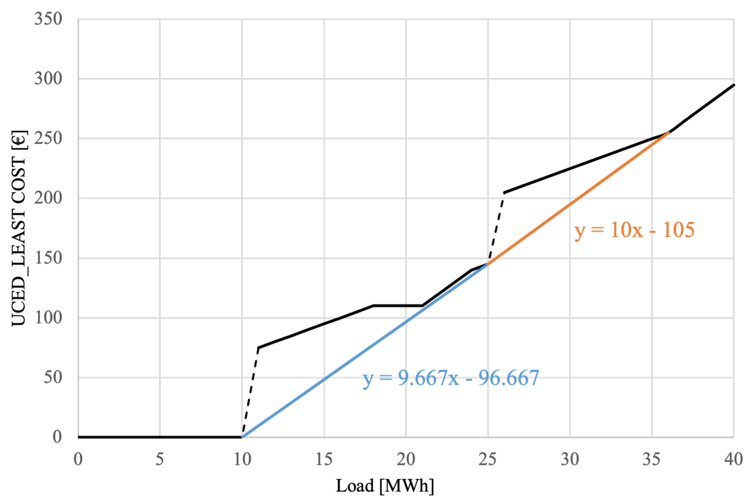


Figure 2.3: Value function $UCED(d)$ (black lines) and its convex hull, as a function of the load d. The convex hull is represented by the black line for $1 \leq d \leq 10$, the blue and the orange lines for $10 \leq d \leq 36$, with slopes 9.667 and 10 respectively, and the black line for $36 \leq d \leq 40$.

$UCED(d)$ at the demand d is the **electricity (CH) price**.

The mathematical counterpart to the graphical approach, suitable for more realistic and complex instances, is known as *Convex Hull – partial Lagrangian* (or *minimum uplift*) and was introduced by [Gribik et al., 2007]. The **CONVEX HULL** (CH) PRICING can be defined as the max-min problem (2.6)

$$\mathcal{L}^* = \max_{\lambda} \hat{\mathcal{L}}(\lambda) = \max_{\lambda} \left[\min_{Q_k, \gamma_k \in \mathcal{X}_k} \mathcal{L}(Q_k, \gamma_k, \lambda) \right] \tag{2.6}$$

obtained from the UCED problem (2.1). In order to solve (2.6), the partial Lagrangian is preliminarily defined as in (2.7)

$$\mathcal{L}(Q_k, \gamma_k, \lambda) = \sum_k (C_k^F \gamma_k + C_k^M Q_k) + \lambda(d - \sum_k Q_k) \tag{2.7}$$

by relaxing the equality balance constraint (2.1b) as a penalty function while introducing the - unrestricted in sign - Lagrange multiplier λ, then proceeds accordingly with Algorithm (1).

Algorithm 1: Convex Hull Pricing

Data: An instance of the clearing problem
Result: $\lambda^\star$
begin
  `Choose` λ
  repeat
    $\hat{\mathcal{L}}(\lambda) \leftarrow$ minimizing $\mathcal{L}(Q_k, \gamma_k, \lambda)$ over $Q_k, \gamma_k \in \mathcal{X}_k$
    $\lambda \leftarrow$ `UpdateLamba`(λ)
  until $\lambda = \lambda^\star$ *s.t.* $\hat{\mathcal{L}}(\lambda) = \max_\lambda \hat{\mathcal{L}}(\lambda)$

The feasible set $\mathcal{X}_k$ is defined by the local constraints for each generator k, depending on the complexity required, e.g. as the ones described in [Bacci et al., 2023] where it is also proposed a specialized algorithm for solving the so-called 1UC, i.e. minimizing $\mathcal{L}(Q_k, \gamma_k, \lambda)$.

The optimal value of the outer variable λ is the **electricity (CH) price**:

$$\lambda^{CH} = argmax_\lambda \hat{\mathcal{L}}(\lambda)$$

. It partially incorporates non-convex cost components, thereby mitigating economic losses. It also reflects part of the marginal cost of the generators dispatched at the technical minimum. In [Pablo Luna et al., 2021] the authors show that the CH problem, as a byproduct, minimizes the difference between the preferred profits $\sum_k \pi_k^{PM}(\lambda)$ and the total payment λd, thus minimizing the LOC. In fact, (2.6) is equivalent to (2.8)

$$\max_\lambda \lambda d + \sum_k \min_{Q_k, \gamma_k \in \mathcal{X}_k} \mathcal{L}_k(Q_k, \gamma_k, \lambda) \tag{2.8}$$

as the Lagrangian function $\mathcal{L}(Q_k, \gamma_k, \lambda)$ is separable into $\mathcal{L}_k = C_k^F \gamma_k + C_k^M Q_k - \lambda Q_k$. Problem (2.8) is in turn equivalent to (2.9)

$$\min_\lambda \sum_k \pi_k^{PM}(\lambda) - \lambda d \tag{2.9}$$

as $\pi_k^{PM} = \max_{Q_k,\gamma_k \in \mathcal{X}_k} \lambda Q_k - C_k^F \gamma_k - C_k^M Q_k$. Due to the balance constraint (2.1b) in the UCED problem, it follows that $d = \sum_k Q_k^{UCED}$. Therefore, (2.9) becomes (2.10)

$$\min_{\lambda} \sum_k \pi_k^{PM}(\lambda) - \lambda \sum_k Q_k^{UCED} \tag{2.10}$$

Adding the constant $\sum_k (C_k^F \gamma_k^{UCED} - C_k^M Q_k^{UCED})$ does not change the optimal value of λ:

$$\begin{aligned}
&\min_{\lambda} \sum_k \pi_k^{PM}(\lambda) - [\lambda \sum_k Q_k^{UCED} - \sum_k (C_k^F \gamma_k^{UCED} - C_k^M Q_k^{UCED})] = \\
&= \min_{\lambda} \sum_k \pi_k^{PM}(\lambda) - \sum_k \pi_k^{UCED}(\lambda) \\
&= \min_{\lambda} \sum_k LOC_k(\lambda)
\end{aligned}$$

Consequently, λ^{CH} minimizes the LOC, thereby necessitating minimal side payments to support the least-cost dispatch.

Obtaining optimal values for λ^{CH} could be computationally expensive as the CH problem requires solving a sub-problem for each generator and at the same time optimizing over a non-differentiable convex function w.r.t λ^{CH}, e.g. [Frangioni, 2005], [Borghetti et al., 2003], possibly requiring specialized approaches such as bundle methods, e.g. [Frangioni, 2002]. In [Andrianesis et al., 2022] it is proposed a Dantzig-Wolfe decomposition approach. To deal with computational problems, the CHPri (2.11), PD (2.12), and LD (2.13) approaches have been proposed, and are described in the following sections.

2.1.1.4 Convex Hull Primal

In [Hua and Baldick, 2017] the authors propose a polynomially-solvable primal formulation for the Lagrangian dual problem for the CH problem, called *Convex Hull Primal*. It is also referred to as *Tight Dispatchable* in [Eldridge et al., 2020].

The **CONVEX HULL PRIMAL** (CHPri) problem is a tight formulation of the UCED problem (2.1) obtained by finding for each generator k the convex envelope of the individual cost function $C_k = C_k^F \gamma_k + C_k^M Q_k$ and the convex hull of the feasible set $\mathcal{X}_k$, defined by local constraints. Therefore, instead of finding the convex envelope of the value function (system cost function), the approach performs a convexification on a resource-specific level. The convex hull $conv(\mathcal{X}_k)$ is represented by all convex combinations of points in $\mathcal{X}_k$ and it is given by the intersection of all convex sets that contain $\mathcal{X}_k$, while the convex envelope is the largest convex function on $conv(\mathcal{X}_k)$ that is an under-estimator of C_k on $\mathcal{X}_k$. The reformulation becomes relevant when start-up/shut-down capabilities and min up/down times are considered. Furthermore, with a single marginal cost per generator, the convex envelope has the same functional form as C_k, since the latter is affine, and only the convex hull needs to be determined. Applying the convex hull description of the UCED problem of [Gentile et al., 2017], the CHPri problem is as in (2.11)

$$\min_{Q_k,\gamma_k} \quad \sum_k [C_k^F \gamma_k + C_k^M (\underline{Q}_k \gamma_k + Q_k)] \tag{2.11a}$$

$$\text{s.t.} \quad \sum_k (\underline{Q}_k \gamma_k + Q_k) = d \qquad \lambda^{CHP} \tag{2.11b}$$

$$0 \leq Q_k \leq (\overline{Q}_k - \underline{Q}_k)\gamma_k \qquad k \in K \tag{2.11c}$$

$$0 \leq \gamma_k \leq 1 \qquad k \in K \tag{2.11d}$$

In (2.11), Q_k represents the power production above the generator's minimum output $\underline{Q}_k$. The optimal value of dual variable λ^{CHP} associated with the supply-demand balance constraint (2.11b) is the **electricity (CHPri) price**.

The CHPri method does not replicate the least-cost dispatch ($\underline{Q}_k \gamma_k^{CHP} + Q_k^{CHP} \neq Q_k^{UCED}$, $\gamma_k^{CHP} \neq \gamma_k^{UCED}$). The price λ^{CHP} can, to some extent, reflect fixed costs and/or the costs of operating at the minimum output of any generator, as γ_k is a continuous variable defined between 0 and 1. [Byers and Eldridge, 2022] shows that after strategic bidding the cost to consumers is higher with R pricing (2.4) than with CH pricing, which is more expensive without strategic bidding. Despite its potential benefits, also the CHPri approach can present computational difficulties and has several counter-intuitive economic properties [Mays et al., 2021].

2.1.1.5 Partial Dispatchable

The approach *Partial Dispatchable* is illustrated in [Eldridge et al., 2020] and it is also referred to by the names *Restricted Convex Hull* in [Eldridge et al., 2018] and *Partial Approximate Convex Hull* in [Byers and Hug, 2023].

After solving the UCED problem (2.1) to find the least-cost dispatch ($Q_k^{UCED}, \gamma_k^{UCED}\ \forall k$), the **PARTIAL DISPATCHABLE MODEL** (PD) (2.12) is considered:

$$\min_{Q_k,\gamma_k} \quad \sum_k (C_k^F \gamma_k + C_k^M Q_k) \tag{2.12a}$$

$$\text{s.t.} \quad \sum_k Q_k = d \qquad \lambda^{PD} \tag{2.12b}$$

$$\underline{Q}_k \gamma_k \leq Q_k \leq \overline{Q}_k \gamma_k \qquad k \in K \tag{2.12c}$$

$$0 \leq \gamma_k \leq \gamma_k^{UCED} \qquad k \in K \tag{2.12d}$$

(2.12) is the LP obtained from the UCED problem by adding constraints (2.12d) that limit the status variables γ_k, re-defined as real non-negative, to take values between 0 and the optimal statuses γ_k^{UCED}. The optimal value of dual variable λ^{PD} associated with the supply-demand balance constraint (2.12b) is the **electricity (PD) price**.

PD does not replicate the least-cost dispatch ($Q_k^{PD} \neq Q_k^{UCED}$, $\gamma_k^{PD} \neq \gamma_k^{UCED}$). The price λ^{PD} partially reflects fixed costs and/or the costs associated with operating at the minimum output. Prices are more stable compared to those obtained with the R approach (2.4) but still exhibit non-monotonic increases with demand.

2.1.1.6 Loose Dispatchable

The *Loose Dispatchable* approach is illustrated in [Eldridge et al., 2020] and it is equivalent to the approach called *Dispatchable* in [Gribik et al., 2007] or *Integer relaxation* in [Herrero et al., 2020]. It is also referred to as *Approximated Convex Hull Pricing* in [Eldridge et al., 2018] because, with a single marginal cost per unit and in the absence of binding ramping constraints, it is equivalent to the CH (2.6) and CHP (2.11) approaches. The **LOOSE DISPATCHABLE MODEL** (LD) (2.13)

$$\min_{Q_k,\gamma_k} \quad \sum_k (C_k^F \gamma_k + C_k^M Q_k) \tag{2.13a}$$

$$\text{s.t.} \quad \sum_k Q_k = d \qquad \lambda^{LD} \tag{2.13b}$$

$$\underline{Q}_k \gamma_k \leq Q_k \leq \overline{Q}_k \gamma_k \qquad k \in K \tag{2.13c}$$

$$0 \leq \gamma_k \leq 1 \qquad k \in K \tag{2.13d}$$

is the LP problem obtained from the UCED problem by adding constraints (2.13d) that limit the status variables γ_k, re-defined as real non-negative, to take values between 0 and 1. The optimal value of dual variable λ^{LD} associated with the supply-demand balance constraint (2.13b) is the **electricity (LD) price** and it represents the average cost at full capacity from the most expensive generator engaged in the LD solution. In fact, (2.13) can be equivalently formulated as the problem (2.14), where $\bar{C_k^M} = C_k^M + \frac{C_k^F}{\overline{Q}_k}$:

$$
\begin{aligned}
\min_{Q_k,\gamma_k} \quad & \sum_k \bar{C_k^M} Q_k && && (2.14a)\\
\text{s.t.} \quad & \sum_k Q_k = d && \lambda^{PD} && (2.14b)\\
& \underline{Q}_k \gamma_k \leq Q_k \leq \overline{Q}_k \gamma_k && k \in K && (2.14c)\\
& 0 \leq \gamma_k \leq 1 && k \in K && (2.14d)
\end{aligned}
$$

LD does not replicate the least-cost dispatch ($Q_k^{LD} \neq Q_k^{UCED}$, $\gamma_k^{LD} \neq \gamma_k^{UCED}$). Prices λ^{LD} partially reflect fixed costs and/or the costs associated with operating at the minimum output and are increasing in demand.

2.1.1.7 Relaxed Minimum Operation

The approach *Relaxed Minimum Operation* is illustrated in [Byers and Hug, 2023] and it is called *Relaxed LMP* in [Mays et al., 2021] or *EcoMin relaxation* in [Herrero et al., 2020].

After solving the UCED problem (2.1) to find the least-cost dispatch ($Q_k^{UCED}, \gamma_k^{UCED}\ \forall k$), the **RELAXED MINIMUM OPERATION MODEL** (RQmin) (2.15) is considered:

$$
\begin{aligned}
\min_{Q_k} \quad & \sum_k (C_k^F \gamma_k^{UCED} + C_k^M Q_k) && && (2.15a)\\
\text{s.t.} \quad & \sum_k Q_k = d && \lambda^{RQm} && (2.15b)\\
& 0 \leq Q_k \leq \overline{Q}_k \gamma_k^{UCED} && k \in K && (2.15c)
\end{aligned}
$$

Model (2.15) is the LP obtained from the UCED problem by replacing γ_k with the optimal states γ_k^{UCED} and relaxing the minimum operating level $\underline{Q}_k$ of the online generators or a subset of generators. When the relaxation involves only fast-start generators, the pricing scheme is called *Fast-start pricing*. The optimal value of the dual variable λ^{RQm} associated with the supply-demand balance constraint (2.15b) is the **electricity (RQmin) price**. This price does not increase monotonically with demand and does not incorporate fixed costs; instead, it reflects the marginal costs of online generators operating at their minimum production levels (according to UCED), even if the least-cost dispatch is not replicated. Given that the price can be established by generators dispatched at the technical minimum, RQmin may create an incentive for other generators, whose production has been reduced to meet the minimum requirements, to increase their output. These generators may adjust their declared minimum power output to their full capacity to "chase" the lost opportunity, causing the system operator to fail in achieving the most efficient commitment in terms of total cost dispatch [Coutu and White, 2014b].

The following pricing schemes (2.16) and (2.26) have been introduced with the objective of achieving **short-term revenue adequacy**. It's worth noting, however, that some of these schemes deviate from the principle (3) of simplicity in understanding. Instead of considering the electricity price as the dual variable of the balance constraint, the RAPA (2.20), DPA (2.21), and LC (2.23) approaches consider a price target or the price as a primal variable that must be close to the target.

2.1.1.8 Average Incremental Cost

The approach *Average Incremental Cost* is illustrated in [Hua and Baldick, 2017], [Liberopoulos and Andrianesis, 2016].

After solving the UCED problem (2.1) to find the least-cost dispatch ($Q_k^{UCED}, \gamma_k^{UCED}\ \forall k$), the **AVERAGE INCREMENTAL COST MODEL** (AIC) (2.16) is considered:

$$\min_{Q_k,\gamma_k} \quad \sum_k (\tilde{C_k^F} \gamma_k + C_k^M Q_k) \tag{2.16a}$$

$$\text{s.t.} \quad \sum_k Q_k = d \qquad \lambda^{AIC} \tag{2.16b}$$

$$\underline{Q}_k \gamma_k \leq Q_k \leq \overline{Q}_k \gamma_k \qquad k \in K \tag{2.16c}$$

$$0 \leq \gamma_k \leq \gamma_k^{UCED} \qquad k \in K \tag{2.16d}$$

(2.16) is the LP obtained from the UCED problem by adding constraints (2.16d) that limit the status variables γ_k, re-defined as non-negative, to take values between 0 and the optimal statuses γ_k^{UCED} and by amortizing the fixed cost C_k^F (of online generators) over the optimal production Q_k^{UCED}: $\tilde{C_k^F} = \frac{C_k^F \overline{Q}_k}{Q_k^{UCED}}$. This reflects a quantity discount obtained when the generators with fixed costs are dispatched at a higher level. The optimal value of the dual variable λ^{AIC} associated with the supply-demand balance constraint (2.16b) is the **electricity (AIC) price**, which can be equivalently determined by the problem

(2.17) [Liberopoulos and Andrianesis, 2016]:

$$\min_{\lambda^{AIC}} \quad \lambda^{AIC} \tag{2.17a}$$

$$\text{s.t.} \quad \lambda^{AIC} Q_k^{UCED} \geq C_k^F \gamma_k^{UCED} + C_k^M Q_k^{UCED} \qquad k \in K \tag{2.17b}$$

For each (online) generator k, the revenue adequacy constraint (2.17b) ensures that the market-based payment $\lambda^{AIC} Q_k^{UCED}$ is sufficient to cover marginal and fixed costs. (2.17) is in turn equivalent to (2.18)

$$\lambda^{AIC} \geq \frac{C_k^F \gamma_k^{UCED} + C_k^M Q_k^{UCED}}{Q_k^{UCED}} \qquad k \in K \tag{2.18}$$

Therefore, the price λ^{AIC} represents both the maximum average cost of the dispatched generators ($\gamma_k^{UCED} = 1$) and the smallest revenue-adequate price (in the short term) under the optimal allocation Q_k^{UCED}. The uniform prices λ^{AIC} are higher than the prices obtained with the previous methods to avoid economic losses. The MWP are embedded in the market-based payments received by all dispatched units, increasing the total payment for the consumers; by reflecting the fixed cost of one generator, the price increases the profits of other generators, leading to a potentially high LOC.

2.1.1.9 Semi-Lagrangian Relaxation

The *Semi-Lagrangian relaxation* approach has been proposed in [Araoz and Jörnsten, 2011]. The **SEMI LAGRANGIAN RELAXATION** (SLR) is the min model (2.19) obtained from the UCED problem (2.1) by semi-relaxing the equality balance constraint (2.1b):

$$\hat{\mathcal{L}}(\lambda) = \min_{Q_k, \gamma_k \in \mathcal{X}_k} \quad \mathcal{L}(Q_k, \gamma_k, \lambda) = \tag{2.19a}$$

$$= \min_{Q_k, \gamma_k \in \mathcal{X}_k} \quad \sum_k (C_k^F \gamma_k + C_k^M Q_k) + \lambda (d - \sum_k Q_k) \tag{2.19b}$$

$$\text{s.t.} \quad \sum_k Q_k \leq d \tag{2.19c}$$

Considering problem (2.19), the partial Lagrangian with respect to the balance constraint, now expressed in the form of the inequality (2.19c), has been formed (2.19b). The Lagrangian multiplier λ prices in the objective function (2.19b) the amount of demand not satisfied: a sufficiently large value of λ ensures that the optimal solution of SLR respects the original balance equality constraint, since the term $(d - \sum_k Q_k)$ in the objective function can only take non-negative values, it can only assume a null value in order to minimize the objective function. This optimal value for λ, i.e., the **electricity (SLR) price** λ^{SLR}, represents the smallest revenue-adequate price (in the short term) for self-interested suppliers [Liberopoulos and Andrianesis, 2016]. With respect to CH (2.6), the SLR problem (2.19) is not separable as the balance coupling constraint is maintained in the inequality form, creating a possible drawback for applying the pricing to large systems since the approach needs to solve a certain number of these problems [Pablo Luna et al., 2021]. The SLR problem is solved iteratively through the Algorithm (2). As for the

Algorithm 2: Semi-Lagrangian Relaxation

Data: An instance of the clearing problem
Result: $\lambda^{\star}$

begin
  $UCED(d) \leftarrow$ `Solve` (2.1) — `// find the minimum cost` $UCED(d)$
  `Choose initial` (feasible) $\lambda = \lambda^1$
  repeat
    $\hat{\mathcal{L}}(\lambda) \leftarrow$ `Solve` (2.19)
    $\lambda \leftarrow$ `IncreaseLambda`(λ)
  until $\lambda = \lambda^{\star}$ *s.t.* $\hat{\mathcal{L}}(\lambda^{\star}) = UCED(d)$

initial price λ^1, one of the non-revenue-adequate prices obtained with the previous approaches can be considered; in [Araoz and Jörnsten, 2011] it is suggested using $\lambda^1 = \lambda^{LD}$, where λ^{LD} is the price obtained with (2.13).

Similar to the AIC approach (2.16), the uniform price λ^{SLR} is received by all dispatched generators; the effect on consumer payments could be even worse, since $\lambda^{SLR} \geq \lambda^{AIC}$.

2.1.1.10 Revenue Adequate Price Adders

The same results of the SLR approach (2.1.1.9) can be obtained by implementing the approach *Revenue Adequate Price Adders* proposed in [Byers and Hug, 2023]. Again, after solving the UCED problem (2.1) to find the least-cost UC $(Q_k^{UCED}, \gamma_k^{UCED}\ \forall k)$, the **REVENUE ADEQUATE PRICE ADDERS MODEL** (RAPA) (2.20) is considered:

$$\min_{\varepsilon} \quad \sum_k \varepsilon Q_k^{UCED} \tag{2.20a}$$

$$\text{s.t.} \quad (\lambda^* + \varepsilon) Q_k^{UCED} \geq C_k^F \gamma_k^{UCED} + C_k^M Q_k^{UCED} \qquad k \in K \tag{2.20b}$$

$$\varepsilon \geq 0 \qquad k \in K \tag{2.20c}$$

Problem (2.20) is a LP and determines the optimal values for the **price adder** ε^{RAPA} so as to satisfy (2.20b), which ensures revenue adequacy (in the short term) for each generator k, and to minimize the total additional payments $\sum_k \varepsilon Q_k^{UCED}$. As the base price λ^*, one of the non-revenue-adequate prices obtained with the previous approaches can be considered.

2.1.1.11 Dual Pricing Algorithm

The *Dual Pricing Algorithm* was proposed in [O'Neill et al., 2017] and later extended in [Hytowitz, 2018]; while both works show a price-responsive demand, [Pablo Luna et al., 2021] illustrates its adaptation to the inelastic demand case.

After solving the UCED problem (2.1) to find the least-cost dispatch $(Q_k^{UCED}, \gamma_k^{UCED})$ for each generator k, the **DUAL PRICING ALGORITHM MODEL** (DPA) reads as in (2.21):

$$\min_{\lambda, U_k} \quad \sum_k U_k + a\frac{|\lambda - \lambda^{IP}|}{\lambda^{IP}} \tag{2.21a}$$

$$\text{s.t.} \quad \lambda Q_k^{UCED} + U_k \geq C_k^F \gamma_k^{UCED} + C_k^M Q_k^{UCED} \qquad k \in K \tag{2.21b}$$

$$\lambda \geq 0 \tag{2.21c}$$

$$U_k \geq 0 \qquad k \in K \tag{2.21d}$$

Model (2.21) is a nonlinear problem (NLP) that determines optimal values for **electricity (DPA) price** λ^{DPA}, which is defined as a primal variable, plus discriminatory **compensations** $U_k^{DPA}, \forall k \in K$. The objective is to satisfy (2.21b), which ensures revenue adequacy (in the short term) for each generator k, and to minimize the sum of the total compensation $\sum_k U_k$ and the distance to the target price, which is assumed to be the price λ^{IP} obtained with (2.5). The scalar a is chosen by the modeler and balances the trade-off between proximity to the target price λ^{IP} (market operator's perspective) and direct payments only (generators' perspective). The absolute value in the objective function can be linearized by introducing an auxiliary variable x as in (2.22):

$$\min_{\lambda, U_k, x} \quad \sum_k U_k + ax \tag{2.22a}$$

$$\text{s.t.} \quad \lambda Q_k^{UCED} + U_k \geq C_k^F \gamma_k^{UCED} + C_k^M Q_k^{UCED} \qquad k \in K \tag{2.22b}$$

$$\lambda \geq 0 \tag{2.22c}$$

$$U_k \geq 0 \qquad k \in K \tag{2.22d}$$

$$x \geq \frac{\lambda - \lambda^{IP}}{\lambda^{IP}} \tag{2.22e}$$

$$x \geq \frac{\lambda^{IP} - \lambda}{\lambda^{IP}} \tag{2.22f}$$

In contrast to the IP approach (2.5), where market-based payments are expressed as $\lambda^{IP} Q_k^{UCED} + \mu_k^{IP} \gamma_k^{UCED}$, the compensation U_k^{DPA} introduced by (2.21) does not depend on the status of generator k. This means that generators might be motivated to deviate from the least-cost solution since they can receive compensation even if they do not operate the plant.

2.1.1.12 Limited Compensation

To limit the compensations received by generators, the *Limited Compensation* approach has been proposed by [Pablo Luna et al., 2021].

After solving the UCED problem (2.1) to find the least-cost dispatch ($Q_k^{UCED}, \gamma_k^{UCED}\ \forall k$), the **LIMITED COMPENSATION MODEL** (LC) (2.23) is considered:

$$\min_{\lambda, U_k, s_k} \quad \frac{1}{2} \|\lambda - \lambda^*\|_2^2 \tag{2.23a}$$

$$\text{s.t.} \quad U_k \leq C_k^F \gamma_k^{UCED} s_k \qquad k \in K \tag{2.23b}$$

$$\lambda Q_k^{UCED} + U_k \geq C_k^F \gamma_k^{UCED} + C_k^M Q_k^{UCED} \qquad k \in K \tag{2.23c}$$

$$\lambda Q_k^{UCED} + U_k \leq C_k^F \gamma_k^{UCED} + C_k^M Q_k^{UCED} + M(1 - s_k) \qquad k \in K \tag{2.23d}$$

$$\sum_k U_k \leq \beta \left[\lambda \sum_k Q_k^{UCED} - \sum_k (C_k^F \gamma_k^{UCED} + C_k^M Q_k^{UCED}) \right] \tag{2.23e}$$

$$\lambda \geq 0 \tag{2.23f}$$

$$U_k \geq 0 \qquad k \in K \tag{2.23g}$$

$$s_k \in \{0, 1\} \qquad k \in K \tag{2.23h}$$

Model (2.23) is a mixed-integer quadratic problem (MIQP) that determines optimal values for **electricity (LC) price** λ^{LC} and discriminatory **compensation** $U_k^{LC}, \forall k \in K$. In the objective function (2.23a), the squared Euclidean distance to the target price λ^* is minimized to stabilize the output and avoid price volatility. Different constraints are considered: (2.23b) limits the compensation to cover fixed costs only, while (2.23c) ensures the revenue adequacy (in the short term) for each generator k. The constrain (2.23d) avoids overcompensation: if $s_k = 1$, the total payment $\lambda Q_k^{UCED} + U_k$ can at most cover the costs; if $s_k = 0$, then $U_k = 0$ due to (2.23b) and the constraint is redundant since M is a large positive constant. Finally, (2.23e) constrains the total compensation $\sum_k U_k$ not to exceed a fraction β of the market profit: when binding, it transfers part of the side payments to the electricity price λ. The modeler chooses the target price λ^* from the non-revenue-adequate prices obtained by the approaches presented above and according to the desirable properties he wishes to maintain. Additionally, he selects the scalar β to limit the total compensation. The compensation U_k^{LC} is always zero for generators with null C_k^F due to (2.23b). Furthermore, the payments cover the marginal costs of plants operating at minimum output thanks to (2.23c). The additional constraint $\sum_k U_k = 0$ can be added but it can result in large increases of the price λ^{LC}.

A simplified version of (2.23) has been proposed in [Byers and Hug, 2023]: instead of the binary variables s_k, (2.24) introduces a price adder $\boldsymbol{\varepsilon}$ on top of the price λ^*, which is no longer optimized (target price).

$$\min_{U_k, \boldsymbol{\varepsilon}} \quad \sum_k (\boldsymbol{\varepsilon} Q_k^{UCED} + U_k) \tag{2.24a}$$

$$\text{s.t.} \quad (\lambda^* + \boldsymbol{\varepsilon}) Q_k^{UCED} + U_k \geq C_k^F \gamma_k^{UCED} + C_k^M Q_k^{UCED} \qquad k \in K \tag{2.24b}$$

$$\sum_k U_k \leq \beta \sum_k (\lambda^* + \boldsymbol{\varepsilon}) Q_k^{UCED} \tag{2.24c}$$

$$U_k \geq 0 \qquad k \in K \tag{2.24d}$$

$$\boldsymbol{\varepsilon} \geq 0 \tag{2.24e}$$

2.1.1.13 Generalized Uplift

The *Generalized Uplift* is illustrated in [Bouffard and Galiana, 2005]. The name "generalized" refers to a discriminatory payment that is given to all the generators in the market, whether they are scheduled on or off, to enable them not only to cover losses but to make the maximum profit at the least-cost dispatch.

After solving the UCED problem (2.1) to find the least-cost dispatch $(Q_k^{UCED}, \gamma_k^{UCED})$

for each generator k, the **GENERALIZED UPLIFT** (GU) (2.25) is considered:

$$\min_{\lambda, \Delta C_k^F, \Delta C_k^M} \quad \sum_k \left[(\Delta C_k^F \gamma_k^{UCED})^2 + (\Delta C_k^M Q_k^{UCED})^2 \right] \tag{2.25a}$$

$$\text{s.t.} \quad \sum_k (\Delta C_k^F \gamma_k^{UCED} + \Delta C_k^M Q_k^{UCED}) = 0 \tag{2.25b}$$

$$(\lambda, \Delta C_k^F, \Delta C_k^M) \in \Omega(Q_k^{UCED}, \gamma_k^{UCED}) \qquad k \in K \tag{2.25c}$$

Model (2.25) is a quadratic problem (QP) that determines optimal values for the **electricity (GU) price** λ^{GU} and the **uplift function value** $U_k = \Delta C_k^F \gamma_k^{UCED} + \Delta C_k^M Q_k^{UCED}$, $\forall k \in K$.

In the objective function (2.25a) the squared norm of the uplift components ΔC_k^F (the component associated with the ON state) and ΔC_k^M (the component associated with the production) is minimized. Constraint (2.25b) ensures the system operator's neutrality through zero-sum transfer between generators, while constraint (2.25c) guarantees that the least-cost solution $(Q_k^{UCED}, \gamma_k^{UCED})$ is the profit-maximizing solution, so that generators do not incur in LOC by following it. In fact, $\Omega(Q_k^{UCED}, \gamma_k^{UCED})$ is the set defining the profit-optimality conditions for the generator k:

$$\Omega(Q_k^{UCED}, \gamma_k^{UCED}) =$$
$$= \{(\lambda, \Delta C_k^F, \Delta C_k^M) : \pi_k(Q_k, \gamma_k) \le \pi_k(Q_k^{UCED}, \gamma_k^{UCED}) - \boldsymbol{\varepsilon}, \forall (Q_k, \gamma_k) \in \mathcal{X}_k\}$$

$\boldsymbol{\varepsilon}$ is a small positive scalar used to guarantee that the least-cost solution $(Q_k^{UCED}, \gamma_k^{UCED})$ leads to the maximum profit in the case of multiple profit-maximizing schedules. Additional constraints could be added, such as those on the rules for sharing the cost to achieve null LOC between generators and consumers. The sharing would be justified by the incentives for generators not to try to artificially cause sub-optimality by their bidding strategies.

The sets $\Omega(Q_k^{UCED}, \gamma_k^{UCED})$ are non-convex and impossible to express explicitly due to the infinite possible feasible dispatches (Q_k, γ_k). For single-time period cases, [Motto and Galiana, 2002] and [Galiana et al., 2003] reformulate each non-convex set into two equivalent linear constraints. However, for multi-period models deriving them is impractical as the number of necessary conditions increases exponentially with the number of time periods. In addressing this challenge, [Bouffard and Galiana, 2005] proposed an iterative method exemplified in Algorithm (3) that sequentially limits the feasible set, by adding user cuts, until the desired conditions are met.

Algorithm 3: Generalized Uplift

Data: An instance of the clearing problem

Result: $\lambda^{\star}$,$\Delta C_k^{F\star}$, $\Delta C_k^{M\star}$

begin

  Set $i = 0$; Set $\Omega_k^{[0]} = \{\emptyset\}$; Set ε

  Set $\Delta C_k^{F[0]} = 0$

  Set $\Delta C_k^{M[0]} = 0$

  $(Q_k^{UCED}, \gamma_k^{UCED}) \leftarrow$ Solve (2.1) // find the least-cost dispatch

  $\lambda^R \leftarrow$ Solve (2.4) // find the restricted price

  Set $\lambda^{[0]} = \lambda^R$

  for $i = 1$ **to** N **do**

    $\forall k : (Q_k^{[i+1]}, \gamma_k^{[i+1]}) \leftarrow$ Solve (2.2) // find the profit-maximizing dispatch considering price $\lambda^{[i]}$ and additional compensations $\Delta C_k^{F[i]}, \Delta C_k^{M[i]}$

    if $(Q_k^{[i+1]}, \gamma_k^{[i+1]}) \neq (Q_k^{UCED}, \gamma_k^{UCED})$ **then**

      $\Omega_k^{[i+1]} = \Omega_k^{[i]} \cap \{(\lambda, \Delta C_k^F, \Delta C_k^M) : \pi_k(Q_k^{[i+1]}, \gamma_k^{[i+1]}) \leq \pi_k(Q_k^{UCED}, \gamma_k^{UCED}) - \varepsilon$ // AddCut to $\Omega_k^{[i+1]}$

    if $\Omega_k^{[i+1]} = \Omega_k^{[i]}$ $\forall k$ **then**

      $\lambda = \lambda^{\star}$

      $\Delta C_k^{F\star} = \Delta C_k^{F[i]}$

      $\Delta C_k^{M\star} = \Delta C_k^{M[i]}$

      STOP // Alg terminate

    else

    $\lambda^{[i+1]}, \Delta C_k^{F[i+1]}, \Delta C_k^{M[i+1]} \leftarrow$ Solve (2.25) // with $\Omega_k^{[i+1]}$ in (2.25c)

    $i \leftarrow i + 1$

2.1.1.14 Primal-Dual

The *Primal-Dual* approach has been proposed in [Ruiz et al., 2012]. The **PRIMAL-DUAL** (2.26) is the mixed-integer bilinear problem (MIBP) obtained by following the steps (1)-(5):

1. considering as the primal problem the one obtained from the UCED problem (2.1) by adding constraints $0 \leq \gamma_k \leq 1, \forall k \in K$ that limit the status variables γ_k, redefined as real non-negative, to take values between 0 and 1;

2. deriving the dual problem of (1);

3. combining primal (1) and dual (2) to formulate a new problem that minimizes the **duality gap** DG (2.26a) and is subject to both primal (2.26b)-(2.26d) and dual (2.26e)-(2.26f) constraints;

4. adding back integrality conditions $\gamma_k \in \{0,1\}, \forall k \in K$ (2.26i);

5. introducing revenue adequacy constraints (2.26j) that ensure that each generator k achieves non-negative profit in the short term.

$$
\begin{aligned}
\min_{Q_k,\gamma_k,\lambda,\underline{\mu}_k,\overline{\mu}_k,\rho_k} \quad & DG = \sum_k (C_k^F \gamma_k + C_k^M Q_k) - (d\lambda - \sum_k \rho_k) && && (2.26a)\\
\text{s.t.} \quad & d - \sum_k Q_k = 0 && && (2.26b)\\
& \underline{Q}_k \gamma_k - Q_k \leq 0 && k \in K && (2.26c)\\
& Q_k - \overline{Q}_k \gamma_k \leq 0 && k \in K && (2.26d)\\
& C_k^M - \lambda - \underline{\mu}_k + \overline{\mu}_k \geq 0 && k \in K && (2.26e)\\
& C_k^F + \underline{Q}_k \underline{\mu}_k - \overline{Q}_k \overline{\mu}_k + \rho_k \geq 0 && k \in K && (2.26f)\\
& \underline{\mu}_k, \overline{\mu}_k, \rho_k \geq 0 && k \in K && (2.26g)\\
& \lambda \textit{ free} && && (2.26h)\\
& \gamma_k \in \{0,1\} && k \in K && (2.26i)\\
& \lambda Q_k \geq C_k^F \gamma_k + C_k^M Q_k && k \in K && (2.26j)
\end{aligned}
$$

In addition to the **electricity (P-D) price** λ^{P-D}, (2.26) determines the optimal values for both **statuses** γ_k^{P-D} and **productions** Q_k^{P-D}, for each generator $k \in K$.

The AIC approach (2.16) determines the minimum revenue-adequate price (in the short term) under the least-cost allocation (γ_k^{UCED}, Q_k^{UCED}), whereas the P-D approach ensures revenue adequacy under the dispatch order (γ_k^{P-D}, Q_k^{P-D}), that may deviate from the least-cost dispatch. For example, P-D may transfer a part of the dispatched quantities and the related payments from the infra-marginal suppliers to the marginal supplier (*cross-subsidy*). This deviation results in a higher cost for the system: P-D trades cost efficiency for price efficiency, as long as this trade-off reduces the duality gap [Liberopoulos and Andrianesis, 2016]. The uniform revenue-adequate price λ^{P-D} affects all generators dispatched, but the effect on consumer payments is smaller ($\lambda^{P-D} \leq \lambda^{AIC}$). The profit π_k^{P-D} is always non-negative thanks to (2.26j), and MWP are not needed. The LOC required to support the P-D dispatch solution can be defined as $LOC_k^{P-D} = \pi_k^{PM} - \pi_k^{P-D}$.

With this definition, the equality (2.27) holds

$$\sum_k LOC_k^{P-D} = DG \tag{2.27}$$

Therefore, by minimizing the duality gap, (2.26) minimizes the LOC associated with the new dispatch order. The CH (2.6) remains the approach that minimizes the LOC associated with the least-cost dispatch.

A slightly different version of (2.26) has been proposed in [Guo et al., 2022]: in the objective function (2.28), instead of considering only the duality gap, both the total cost and the duality gap are minimized, with a regularization term r acting as a tunable parameter between the two objectives.

$$\min_{Q_k,\gamma_k,\lambda,\underline{\mu}_k,\overline{\mu}_k,\rho_k} \sum_k (C_k^F \gamma_k + C_k^M Q_k) + rDG \tag{2.28}$$

The scalar r is chosen by the modeler and balances the trade-off between minimizing system costs and the deviation from the competitive equilibrium.

Unlike the IP approach (2.5), the P-D approach may discourage participants from adopting strategic pay-as-bid , as there is no straightforward relationship between the electricity price and the "declared" costs. However, this could lead to problems of price interpretability.

Tab. 2.3 and Tab. 2.4 summarize the main characteristics of the pricing schemes previously described.

2.1.2 Bidding behavior on the Italian day-ahead market

In general EU electricity markets adopt a different approach in dealing with non-convexities, with the goal of facilitating trading and maximizing transparency (2) [Herrero et al., 2020]. The basic bidding format is the (quantity-price) pairs. However, in some European countries, additional conditions can then be imposed, as the ones illustrated in Tab. 2.5. For example, in the Iberian market, fixed costs are not taken into account in the determination of the least-cost dispatch, but if the *minimum income conditions (MIC)* are not met, the order is not accepted.

The Italian day-ahead market (MGP) considers a mechanism based on simple single-hour bids that enables producers to submit proposals containing only the indication of the offered amount and the requested price. According to [GME, 2016], [GME, 2019], each generator can submit a maximum of 8 non-decreasing in price bids per hour: up to

R	Traditional model used to compute electricity prices in US
IP	Nullify LOC (The modified versions present LOC>0 but allow infra-marginal generators to keep their profits or determine non-discriminatory and non-volatile payments)
CH	Minimize LOC
CHPri	Primal formulation of CH, computationally more tractable
PD	Reflect fixed costs (partially) and marginal costs of generators dispatched at the minimum level
LD	Partially reflect both the fixed and the marginal costs of generators dispatched at the minimum level, without depending on the UCED solution
RQmin	Reflect the marginal costs of generators dispatched at the minimum level
AIC	Find the smallest revenue-adequate price under the least-cost dispatch
SLR	Find the smallest revenue-adequate price for the self-interested generators
RAPA	Find a non-discriminatory adder on price to make the least-cost solution revenue-adequate
DPA	Trade-off between deviation from the IP price and minimization of the discriminatory compensations
LC	Limit the discriminatory compensations
GU	Ensure revenue neutrality for the system operator and zero LOC for generators
P-D	Find both dispatch and revenue-adequate price; the dispatch differs from the least-cost dispatch. Trade-off between cost minimization and competitive equilibrium

Table 2.3: Main goal for each pricing scheme

four supply offers and up to four injection programs related to bilateral contracts. The latter are treated as supply offers to account for the corresponding power flow on the transmission network between zones as defined by the MGP rules. The bidding curve of generator k in hour t is therefore the non-decreasing step-wise linear function obtained by representing quantity-price $(Q_{k,t,s}, P_{k,t,s})$ pairs in non-decreasing order of price, where s indicates the step. In Fig. 2.4, examples of bid curves with different numbers of steps are illustrated. We will detail the mathematical formulation of the Italian MGP model in (4.1.1).

The articulation of the bidding curve in steps enables the operator to adopt different strategies in formulating bids for the available capacity of its generators. A portion of the capacity, generally equal to the technical minimum or higher, either in cases where the operator wishes to fulfill future or bilateral contracts or in the case of green energy that has to be produced [Muñoz et al., 2013], can be offered at zero or at a reduced price (compared

	R	IP	CH, CHPri	PD	LD	RQmin	AIC	SLR	RAPA	DPA	LC	GU	P-D
1	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	N
2	Y	Y	N	N	N	N	N	Y	Y	Y	Y	Y	N
3	N	Y	P	P	P	N	Y	Y	Y	Y	Y	Y	Y
4	N	Y	P	Y	P	Y	Y	Y	Y	Y	Y	Y	Y
5	N	Y	N	N	N	N	N	N	N	Y	Y	Y	N
6	Y	Y	N	P	N	Y	Y	Y	P	Y	P	Y	P
7	> 0	0	min	> 0	> 0	> 0	0	0	0	0	0	0	0*
8	> 0	0	min	> 0	> 0	> 0	> 0	> 0	> 0	> 0	> 0	0	min*
9	LP	LP	$\text{MIP}^{\text{max-min}}$	LP	LP	LP	LP	MIP	LP	NLP	MIQP	QP	MIBP

Table 2.4: Main characteristics for each pricing scheme (N:no, Y:yes, P:partially).
1: UCED (2.1) dependent;
2: UCED dispatch reproduced by PRICING (For all pricing schemes except for P-D, the considered dispatch order is the least cost, even when the PRICING model does not reproduce it);
3: reflects fixed costs;
4: reflects marginal costs when generators are dispatched at the technical minimum;
5: discriminatory;
6: volatile;
7: MWP (If MWP=0, the price is revenue-adequate in the short term);
8: LOC;
9: class.
* only if the P-D dispatch order is considered.

Block orders	Reject the bid if it is not totally accepted in a period of consecutive hours
Minimum income	Reject the bid if its daily remuneration does not cover a fixed cost (representing, for example, the startup cost) and a variable cost
Load gradient	Reject the bid if the difference between the energy in adjacent hours is greater than a given value

Table 2.5: Complex bids in Europe

to the marginal cost of the plant) in order to increase the probability of the offer being accepted. The generator is then paid the market clearing price (presumably higher than the bid price) thanks to the uniform price auction mechanism (*pay-as-cleared*).

The operator can recover the fixed costs of the generator by offering a portion of its capacity at a bidding price defined as the marginal cost plus an appropriate *bid-up*.

The peculiar configuration of the Italian electricity market, characterized by uniform pricing, almost total rigid demand, and market segmentation into zones, together with the operator's diversified plant portfolio by zone and technology, may allow the operator to influence the market price. In fact, the operator can determine an increase in the clearing price by forcing a higher-cost generator to become marginal through two possible

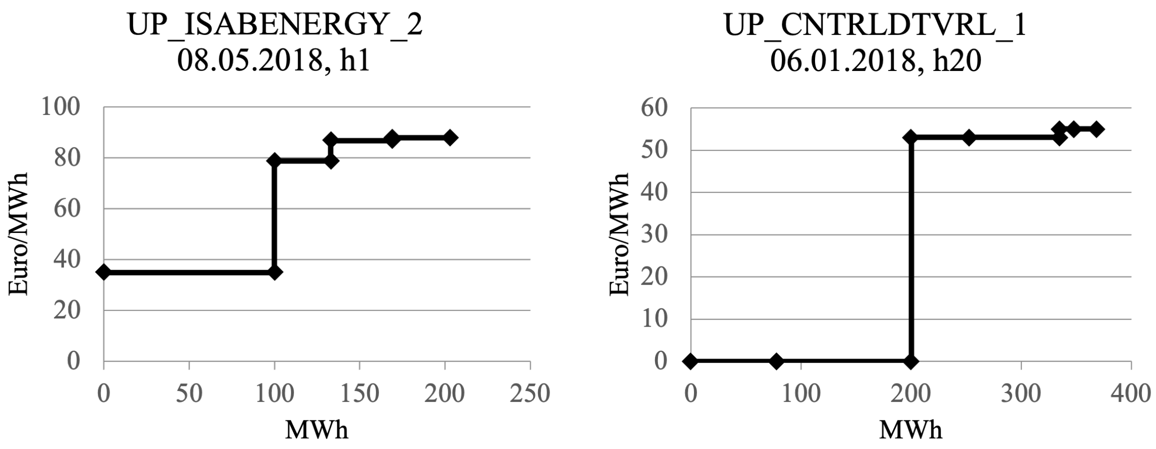


Figure 2.4: Examples of bidding curves with different numbers of steps, showing respectively four and six offers on MGP for the specified generator, hour and day

strategies: inducing congestion between zones and withholding capacity.

The first strategy involves operating even at a loss in one zone in order to congest the transmission lines and isolate the adjacent zone: the generator in the exporting zone bids below its marginal cost to overproduce, forcing a more expensive generator to set the price in the importing zone. The operator's plants in the import zone enjoy a higher price thanks to the network congestion created [Cazzola and Creatini, 2002], [Quick and Carey, 2002].

The generation capacity withholding strategy, on the other hand, refers to the practice of keeping available generation capacity from being competitively offered on the market [García and Reitzes, 2007]. According to the EU Agency for the Cooperation of Energy Regulators (ACER) [ACER, 2021b], the withholding can occur in two ways

1. *economic withholding*

2. *physical withholding*

depending on whether the limited production is determined (1) by offering a quantity at an artificially high price, i.e., a price that does not reflect the marginal cost (including opportunity cost) of the generator and is assumed to be higher than the clearing price, or (2) by not offering the available generation capacity at any price.

The withholding of a significant amount of capacity on the day-ahead and intra-day markets could also be explained by the operator's desire to reserve the available capacity for ancillary services markets, which may be more remunerative. An example of such manipulative behavior by Enel Produzione is described in [AGCM, 2021], where the

Italian Competition Authority (AGCM) reports that some generators, usually dispatched, recorded zero injection programs on MGP+MI during some months of 2016, subsequently forcing Terna to accept its offers on MSD at high prices for ancillary services such as reactive power and grid stability.

When the unit is pivotal, the economic withholding offer might be accepted, leading to a price increase without a reduction in production. In this case, it is more accurate to refer to it as a price-maker strategy: the operator submits offers at a significantly high price, knowing that the production from its generator is essential to meet the demand.

Article 5 of the, constantly evolving, EU Regulation on wholesale Energy Market Integrity and Transparency (REMIT) prohibits market manipulation (or attempted manipulation) of the market but a case-by-case analysis needs to be performed to evaluate the specific violation [ACER, 2021a].

In order to understand how different operators offer the available capacity on the MGP, we perform the following analysis for each hour of the period January 2018 - June 2020. The installed capacity Q_k^I of the generator k is divided into the sum of (1)-(4), each of which can take a positive or zero value at a given hour t, depending on the behavior of the operator in formulating the offer:

1. Priority capacity $Q_{k,t}^{P=0}$ offered at zero price
2. Positive price capacity $Q_{k,t}^{P>0}$, which in turn can be divided into:
   2.1 Market capacity $Q_{k,t}^{0<P\leq P_z^{thr}}$
   2.2 Economically withheld capacity $Q_{k,t}^{P>P_z^{thr}}$
3. Available capacity $Q_{k,t}^{M}$
4. Physically withheld capacity $Q_{k,t}^{Wp} = Q_{k,t}^{M} - \sum_s Q_{k,t,s}$ that is not offered even if available

Operators anticipate the market price, submitting their bids before market clearing. We establish a threshold price P_z^{thr} for each market zone z as the 99th percentile of the distribution of zonal market prices P_z within the considered time period. This threshold reveals the operator's intention to economically withhold capacity, regardless of success. Estimation errors are considered in the 1% of the highest clearing price values. In Fig. 2.5, an example of the partition of the installed capacity is illustrated.

UP_CTE_DEL_M_2 19.10.2019, h1		
s	Q_s	P_s
1	160	0
2	67	33.48
3	67	35.61
4	0.1	499
5	0.1	499
6	0.1	499
7	0.1	499
8	0.1	499

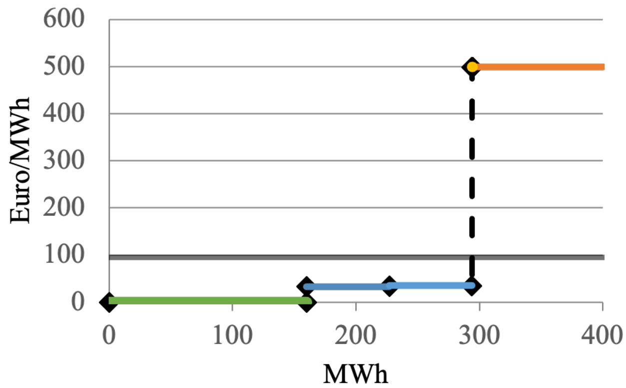


Figure 2.5: Example of the partition of the installed capacity (400MW, equal to the available capacity) between priority capacity (green), market capacity (blue), economically withheld capacity (yellow) and physically withheld capacity (orange), taking into account the offers reported in the table (quantity Q_s, price P_s for each step s offered by the generator on MGP on the specified day and hour) and a threshold of 95 €/MWh (gray) based on the generator's bidding zone

Given the large amount of data (21,888 hours in the period considered), in order to illustrate the strategies adopted by the operator, we represent the distribution of the installed capacity of the generator k at a given hour t as a point with 4 dimensions, or attributes, each defined as a percentage of the installed capacity Q_k^I:

$$(Q_{k,t}^{(P=0)\%}, Q_{k,t}^{(0<P\leq P_z^{thr})\%}, Q_{k,t}^{(P>P_z^{thr})\%}, Q_{k,t}^{Wp\%})$$

Each attribute is defined in the interval [0:1] to make the strategies of different generators comparable and to implicitly take into account the unavailable capacity, represented by $1-(Q_{k,t}^{(P=0)\%}+Q_{k,t}^{(0<P\leq P_z^{thr})\%}+Q_{k,t}^{(P>P_z^{thr})\%}+Q_{k,t}^{Wp\%})$.

We then apply a k-medoids clustering algorithm to the set of points associated with each generator during the hours within the period under consideration.

Clustering is a technique that categorizes the observations of a dataset into groups known as clusters. Specifically, the k-medoids algorithm defines these clusters based on similarities (or dissimilarities) between observations and medoids, which are the representative observations, one per cluster.

For our analysis, we select the *Clustering LARge Applications* (CLARA) algorithm from the k-medoids family, which is specifically designed for large datasets, after determining the optimal number of clusters using the Gap Statistic Algorithm (GSA) [Tibshirani et al., 2001] illustrated by Algorithm (4). For a number of clusters k from one to the number to test k^M (defined by the user), the GSA applies CLARA to find k clusters in both the dataset and the B samples. Each sample is defined considering a reference distribution for each attribute of the data and represents the null hypothesis where no natural clustering structure exists. Then it compares the Total Deviation TD_k, i.e. the distance between the k medoids found with CLARA and the observations of the dataset, to that expected e, determining the value gap_k. The optimal number of clusters k^{opt} is determined by the point where increasing the number of clusters no longer provides a significant improvement in the gap value, also considering the simulation error s_{k+1} with one additional cluster.

Algorithm 4: Gap Statistic Algorithm

Input : Dataset D, max number of clusters k^M, number of bootstrap samples B
Output: Optimal number of clusters k^{opt}

for $k = 1$ *to* k^M **do**
  Apply CLARA to find k clusters in D
  Determine $\ln(TD_k)$

for $b = 1$ *to* B **do**
  Extract a sample S of observations equal to the size of D
  for *each* $k = 1$ *to* k^M **do**
    Apply CLARA to find k clusters in S
    Determine $\ln(TD_{k,b})$

for $k = 1$ *to* k^M **do**
  $e = \frac{1}{B}\sum_{b=1}^{B}\ln(TD_{k,b})$
  $gap_k = e - \ln(TD_k)$
  $sd_k = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left[\ln(TD_{k,b}) - e\right]^2}$
  $s_k = sd_k\sqrt{1 + \frac{1}{B}}$

$k^{\text{opt}} =$ smallest k such that $gap_k \geq gap_{k+1} - s_{k+1}$

For each generator k, we therefore determine the medoids representative of the capacity allocation strategy. An example is shown in Fig. 2.6, where medoid 1 represents the physical withholding strategy of all the available capacity while medoid 4 refers to the cases with unavailability; the percentages of priority capacity $Q^{(P=0)\%}$ may be related to two different plant's configurations or to bilateral contracts while the small percentages of withheld capacity may refer to primary control reserve.

We finally group the medoids by operator in order to compare the strategies adopted on different generators. For example, Tab. 2.6 shows for the generators of A2A the first two medoids by cluster size, excluding the medoids (0,0,0,0) referring to situations of complete unavailability of the plants. The prevailing pattern in the zone NORD indicates that a majority of combined cycle plants adopt economic withholding as their primary

Generator	$\left(Q^{(P=0)\%}, Q^{(0<P\leq P^{thr})\%}, Q^{(P>P^{thr})\%}, Q^{Wp\%}\right)$				Tech	Zone
UP_CASSANO_2	0	0	93	7	CCGT	NORD
UP_CASSANO_2	46	47	0	7	CCGT	NORD
UP_CTE_DEL_M_2	0	0	92	8	CCGT	NORD
UP_CTE_DEL_M_2	0	0	87	13	CCGT	NORD
UP_SERMIDE_3	0	0	95	5	CCGT	NORD
UP_SERMIDE_3	0	0	0	100	CCGT	NORD
UP_SERMIDE_4	0	0	95	5	CCGT	NORD
UP_SERMIDE_4	23	0	0	0	CCGT	NORD
UP_CHIVASSO_1	44	50	0	7	CCGT	NORD
UP_CHIVASSO_1	0	0	96	5	CCGT	NORD
UP_CHIVASSO_2	0	0	95	5	CCGT	NORD
UP_CHIVASSO_2	45	46	0	9	CCGT	NORD
UP_PIACENZA_4	0	0	92	8	CCGT	NORD
UP_PIACENZA_4	0	0	0	50	CCGT	NORD
UP_MONFALCO_1	0	0	97	3	COAL	NORD
UP_MONFALCO_1	80	18	0	3	COAL	NORD
UP_MONFALCO_2	75	21	0	4	COAL	NORD
UP_MONFALCO_2	0	0	96	4	COAL	NORD
UP_GISSI_1	0	0	92	8	CCGT	CSUD
UP_GISSI_1	55	39	0	6	CCGT	CSUD
UP_GISSI_2	0	0	93	7	CCGT	CSUD
UP_GISSI_2	52	39	0	9	CCGT	CSUD
UP_S.F._DEL_1	0	0	0	100	ST	SICI
UP_S.F._DEL_1	0	100	0	0	ST	SICI
UP_S.F._DEL_2	0	0	0	100	ST	SICI
UP_S.F._DEL_2	0	100	0	0	ST	SICI
UP_S.F._DEL_5	0	0	0	100	ST	SICI
UP_S.F._DEL_5	0	100	0	0	ST	SICI
UP_S.F._DEL_6	0	0	0	100	ST	SICI
UP_S.F._DEL_6	0	100	0	0	ST	SICI

Table 2.6: Main medoids for each A2A generator considering the offers on the MGP in the period January 2018 - June 2020: each medoid is characterized by the percentage of capacity offered at zero price, the percentage offered at a market price, and the economically and physically withheld percentages; the colors highlight the non-zero values, from the lowest (yellow) to the highest (red). The table also shows the technology and bidding zone for each generator.

UP_TURBIGO_4					
m	$Q^{(P=0)\%}$	$Q^{(0<P\le P_z^{thr})\%}$	$Q^{(P>P_z^{thr})\%}$	$Q^{Wp\%}$	size
1	0	0	0	1	13167
2	0.48	0.39	0	0.13	2913
3	0.23	0.19	0	0.59	921
4	0	0	0	0	2292
5	0.23	0.64	0	0.14	397
6	0.48	0.45	0	0.07	2198

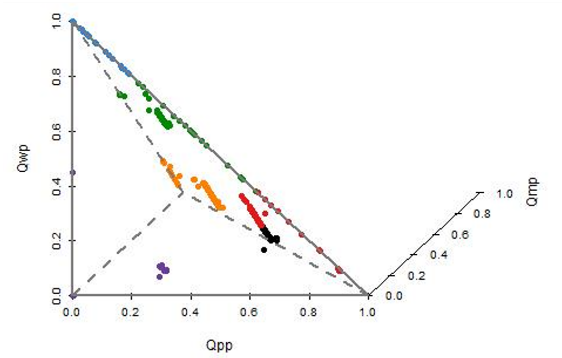


Figure 2.6: Example of the medoids identified by the algorithms for the specified generator; each medoid is characterized by the percentage of capacity offered at zero, offered at market price, economically and physically withheld. The graph shows each observation in terms of priority Qpp, market Qmp, and physically withheld Qwp percentages, as the economically withheld capacity is zero. The colors of the observations indicate the clusters associated with the medoids.

strategy and offer a part of the capacity at zero price and a part of the capacity at market price as a secondary strategy. The coal-fired plants, on the other hand, tend to favor either a zero-price strategy for a substantial portion of their capacity or opt for an economic withholding strategy. For generators located in Sicily, the predominant strategy involves physical withholding for the entire capacity, followed by a market-price bidding strategy, adopted in far fewer cases (< size of the clusters). Small values of $Q^{Wp\%}$ (3%-13%) may suggest the reservation of a capacity for primary reserve purposes rather than a strategy involving physical withholding.

2.2 GEP literature review

Traditional GEP models described in (2.2.1) are quantity-based and lack consideration for prices, thus failing to ensure the economic sustainability of generators. This oversight can lead to loss-making resource mixes for producers, requiring compensatory side payments. Conversely, the works in (2.2.2), along with some references therein, propose different methods for determining a resource mix that takes into account the revenue adequacy of generators.

2.2.1 Traditional GEP models

The decision-making process for planning electric energy systems is characterized by three key features: long-term view, uncertainty, and high dimensionality. Indeed, power generation assets have a lifetime of up to fifty years, so investment decisions have a lasting impact on the future operation of the system and need to take into account the uncertainty arising from, for example, future load or evolution of costs. In order to make informed decisions, large-scale models include many variables and constraints to represent different operating conditions, and in particular integer variables to represent the selection of the facilities.

The seminal book referred to in [Conejo et al., 2016], and the references therein, delve into several aspects of the GEP problems:

1. **Perspective**: centralized or decentralized
2. **Time framework**: static or dynamic
3. **Modeling of the transmission network**: single-node or network-constrained
4. **Modeling of the operating conditions**: historical data or clustering
5. **Unit Commitment Formulations**: binary or clustered
6. **Modeling of the market clearing**: single-level or bilevel
7. **Uncertainty characterization**: deterministic or uncertain

Other challenges, such as the co-optimization of generation and transmission (Generation and Transmission Expansion Planning models - GTEP) or the modeling of interconnections between power systems and other sectors are described in [Koltsaklis and Dagoumas, 2018].

One fundamental aspect is the perspective (1), which can be either centralized or decentralized. In *centralized* GEP models, an independent entity called the Central Planner (CP) performs the *anticipative planning*: given the existing generation portfolio, it determines the long-term resource mix that is most beneficial for the operation of the electric energy system as a whole. With this purpose, the CP can consider different objective functions, such as maximizing social welfare or minimizing different types of costs (investment, operational, and social costs). The CP does not actually build the generating units but uses the outcomes from the analysis as a guide to design different types of policies and incentives (e.g., a price guarantee or capacity payments) to encourage private investors to build certain generation facilities (e.g. those that maximize social welfare). In *decentralized* GEP models, instead, the private profit-oriented endeavor is considered: producers compete with each other to build and operate their respective production facilities with the objective of maximizing profit. Some papers such as [Wogrin et al., 2011] consider a single power producer that competes with other producers in an electricity market: the producer strategically makes investment and operational decisions, exercising market power, while the decisions of its rivals are assumed to be exogenous data. Other works, instead, consider investment equilibrium problems where multiple strategic producers make their own decisions seeking their own profit: at the equilibrium point, no producer can increase its profit by unilaterally changing its strategies. While the decentralized approaches can be used by the individual investor, the centralized ones can be used by the policymakers and Regulators to gain insight into the of strategic producers in response to already specified policies. Centralized models generally allow to consider short-term operations in great detail while do not consider strategic behaviors. In contrast, decentralized approaches are focused on the interactions between strategic players while reducing technical details. Moreover, decentralized GEP may result in discontinuous system development because of inappropriate coordination between actors.

Besides deciding the optimal technology, capacity, and possibly location of the new generating units, it is generally important to determine the optimal timing for building them (2). In this sense, GEP models are classified as *dynamic* (also called multistage), if the evolution of the system is determined over a time horizon consisting of a sequence of reference periods (e.g. years), or as *static* (single-stage), if they take a given future year as a reference. Since the generation expansion plan is mainly conditioned by system demand, which generally increases over time, the reference year is usually chosen as the last year of the planning horizon. If a dynamic framework is considered, then simplifying assumptions in the description of the system is generally required to avoid the problem becoming computationally intractable. A static framework, on the other hand, usually allows for a detailed representation of the system components and can be used with a rolling-window perspective: the static approach can be applied sequentially to different time windows to get an updated view of the situation over time.

Identifying the best locations for new generators is becoming increasingly critical as the penetration of renewables grows, especially in those areas where they are favored for production. This may lead to structural congestion in the transmission network (3). Some studies focus on (*single-node*) systems without representing the transmission network, while others integrate network representation into the GEP problem (*network-constrained*), thus incorporating signals for constructing generating facilities in specific zones and alleviating congestion.

Different operating conditions of the system, such as different demand realizations or capacity factors of renewables, can be modeled in different ways (4): one alternative, often impractical due to the data volume of large systems, consists of considering historical data to predict the operating conditions in the considered planning horizon. To provide a meaningful representation of the system while keeping the problem computationally tractable, some GEP use a small number of *representative periods* (e.g. days) instead of modeling every hour of the planning horizon. There are several techniques for selecting representative days, such as methods based on the load–duration curve and clustering methods [Li et al., 2022]. The latter group similar days into clusters and assign a weight to each cluster so that, for example, the weighted sum of the hourly demand of the clusters is equal to the original demand of the entire planning horizon. The use of representative days prompts the question of how these days should be connected. The most straightforward approach is to assume that all thermal plants are off-line at the beginning of each representative day. However, this results in an overestimation of costs. For a more accurate unit commitment, the initialization approach proposed in [Micheli et al., 2021] can be applied.

In a real test case, incorporating traditional UC constraints (*binary formulation*) for individual power plants would be computationally infeasible due to the number of variables and constraints. Therefore, *clustered unit commitment* formulation can be implemented to maintain the problem computationally tractable (5): similar generators are grouped into a cluster and a single **integer variable** (rather than multiple binary variables) is introduced to represent, for example, the number of online units within each cluster in each time step. Together with the number of variables describing commitment statuses, also the number of constraints and continuous variables representing dispatching decisions is reduced, being defined for only a small number of clusters rather than the complete set of thermal power plants. A comparison of the two formulations is provided in [Meus et al., 2018].

The GEP problem can be modeled as an optimization problem of the protagonist, the CP, subject to the outcomes of the market clearing problem since the market is affected by the generation expansion plan. As a result, the GEP mathematically becomes a *bilevel model* (6), a class of optimization problems that will be illustrated in Chapter (3): the CP receives accepted power quantities from the Market Operator, who, in turn, bases its decisions on the expansion plan.

As mentioned above, GEP models are characterized by long-term planning horizons and this requires to account for uncertainty (7). GEP can be classified as *deterministic* or *uncertain*, depending on whether the planner assumes to have reliable data at the time of decision or makes decisions in an uncertain environment. To deal with uncertainty, different techniques are available, such as stochastic programming or robust optimization [Roald et al., 2023]. In the former case, scenarios are used that model the future realization of uncertain parameters over the planning horizon considered; however, assigning the probability associated with each scenario is challenging, as this information is often unavailable or difficult to estimate. Robust optimization, on the other hand, addresses the uncertain nature of the problem by constraining the uncertain parameters to a deterministic set of values, basically aiming at obtaining solutions that are "good" for any *ex-post* realization of the uncertain data.

GEP models can be formulated as problems that determine the optimal value of decision variables such as new investment, i.e n, dispatched quantities, i.e. Q, and statuses of generators, i.e. γ, while minimizing investment $f^{INV}(n)$ and operational $f^{OP}(Q,\gamma)$ function costs and considering both investment $g^{INV}(n)$ and operational $g(Q,\gamma)$, $h(n,Q,\gamma)$ constraints, such as minimum penetration of renewables or UC constraints. Thus, traditional GEP models are quantity-based and do not take into account pricing, and the relative revenues. After solving this kind of GEP, an *ex-post* analysis can be performed to determine possible losses of selected generators. As illustrated in Fig. 2.7, once the optimal mix has been determined (the investment variables n are fixed to their optimal values n^*), together with the operating conditions (the UC binary variables γ are fixed at their optimal values γ^*), the model becomes linear. Therefore, the clearing prices can be determined as the dual variables λ of the supply-demand balance constraint. At this point, the *Make-Whole Payment* (MWP), i.e., the payment required for the generator k involved to recover both the investment and operational costs, can be eventually calculated as the difference between costs and revenues.

$$MWP_k = -\min\{0, revenues_k - costs_k\} = -\min\{0, \lambda^* \, Q_k^* - f_k^{INV}(n_k^*) - f_k^{OP}(Q_k^*, \gamma_k^*)\}$$

2.2.2 GEP models that account for the Revenue Adequacy

The main GEP works in the literature that consider revenue adequacy are

1. Enumerative approach [Herrero et al., 2015]

2. Heuristic approach [Byers and Hug, 2023]

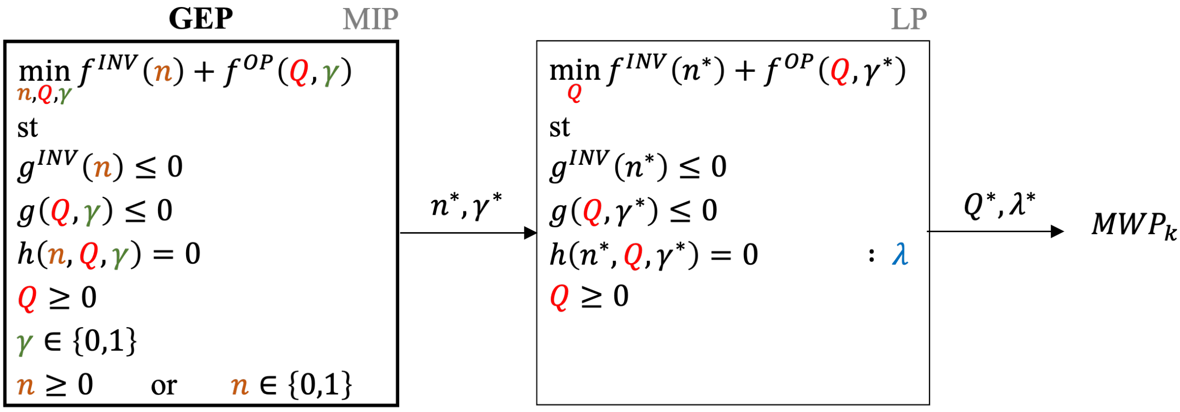


Figure 2.7: Traditional mixed-integer GEP model (in the bold box) and ex-post calculation of the MWP of each generator k: the integer variables, such as investments n and statuses γ, are fixed to their optimal values in order to derive the price λ as the dual variable of the balance constraint in the resulting linear model

3. Optimization approach [Guo et al., 2022]

Note that all consider an energy-only market and a centralized approach.

The *Enumerative approach* (1) investigates how different pricing schemes affect investment decisions by computing long-term profits $\pi_{k,m}^{LT}$ for each technology (or generator) k and for each possible mix m, and then choosing among the mixes with non-negative profits for all generators the one that maximizes the social welfare, as shown in Fig. 2.8. In particular, after solving a traditional GEP to find the least-cost energy mix, it enumerates the set M of all possible mixes, excluding those that differ significantly from the least-cost energy mix. For each energy mix in the set (with the investment decision fixed by the energy mix considered), it determines the least-cost dispatch (2.1) and the clearing price with the R (2.4) (supplemented with MWP to cover losses related to the operational costs) and LD (2.13) pricing schemes. For each pricing scheme and for each technology, the area of all possible combinations is divided into a region of mixes that would make all units of the technology recover their capital cost and a region where not all units of the technology recover their capital costs, defining the break-even frontier of the technology (frontier in the case of two technologies, surface in the case of three technologies). The intersection of the break-even frontiers determines the energy mix chosen for each pricing rule. When investment decisions are discrete, the intersection of the break-even frontiers may not coincide with a combination of power plants; in this case, the mix that maximizes the social benefit is chosen from all quasi-break-even candidates. The investment decision is determined by enumerating all possible solutions, which makes the approach

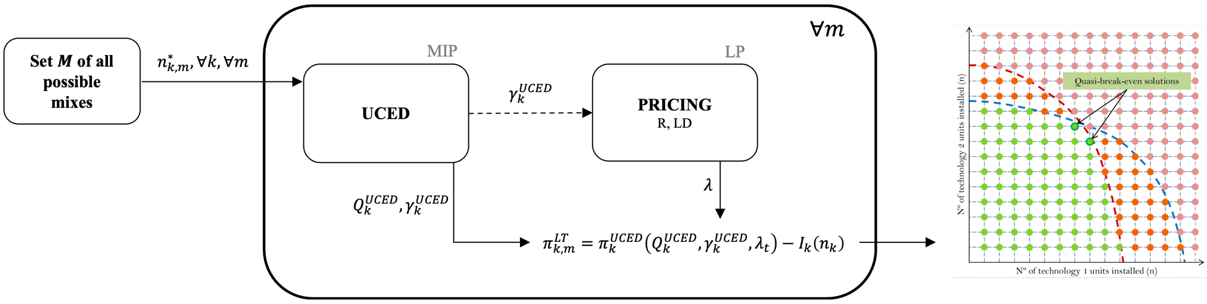


Figure 2.8: Enumerative approach procedure consisting of computing the long-term profits $\pi^{LT}_{k,m}$ for each technology k and possible mix m and selecting the mix that maximizes the social benefit among those with non-negative profit for all technologies; the figure on the right with two technologies and discrete investment decisions is taken from [Herrero et al., 2015]

difficult to apply to large, non-greenfield cases.

The *Heuristic approach* (2) adapts the resource mix to be revenue-adequate for each of the pricing schemes considered. In particular, starting from the solution of a traditional GEP (2.7), for each generator $k \in K$

$$n_k^* = \begin{cases} 0 & \text{if generator } k \text{ is not built in GEP} \\ 1 & \text{if generator } k \text{ is built in GEP} \end{cases}$$

the CP iteratively modifies the investment decisions n_k, $\forall k \in K$ until a *long-term market equilibrium* is reached: considering both the operational costs and the investment costs, no one incurs a loss and no one can make a profit by entering the market. Indeed, both positive and negative profits could result in a move away from the CP solution to an alternate resource mix that may clear a different level of demand. The maximization of the social welfare (or the minimization of the costs, in case of rigid demand) is no longer guaranteed; the method begins with the traditional GEP solution only in the expectation that the equilibrium point is close to the maximum-welfare solution. Since lumpy investments are considered (the investment decision is represented by a binary variable), multiple solutions may exist. Long-term profits π_k^{LT} are determined with the investment decision fixed, following the scheme in Fig. 2.9: given the investments $n_k^{[i]}$ in iteration i, the UCED (2.1), and one of the listed PRICING[4], are solved to determine the dispatch and the clearing price (with any additional compensation).

The investment decisions are iteratively modified according to the Algorithm (5). Given a pricing scheme, e.g. (2.4), the algorithm determines the revenue-adequate re-

[4]The work considers LP PRICING only: R (2.4), IP (2.5), PD (2.12), RQmin (2.15), AIC (2.16), RAPA (2.20) and LC (2.24)

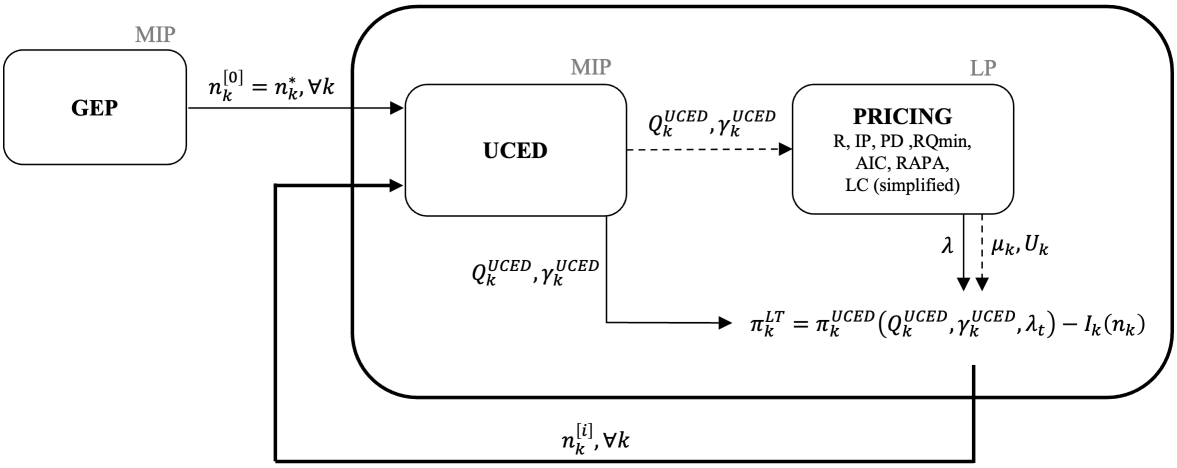


Figure 2.9: Heuristic approach procedure starting from the investment decisions $n_k^{[0]} = n_k^*$ determined by a traditional GEP and consisting of iteratively modifying the investment decisions $n_k^{[i]}$ until the long-term profit π_k^{LT} for each generator k meets certain criteria

source mix $n_k^{RA}, \forall k$. In the initialization part, the set K is ordered to ensure technology neutrality: according to n_k^*, the ordering is done by first considering the built generators and then the unbuilt generators, one per technology. A logic variable $check(g)$ takes the value $TRUE$ if technology g has to be checked, $FALSE$ otherwise. The outer iteration ($while$) is run if for at least one generator k the profit π_k^{LT} is $> bound$ or $< -bound$ and terminates if either the profits of all generators are null (between $-bound$ and $+bound$) or no generator has been added or removed in the last inner iteration ($j = i$). The *bound* instead of 0 is considered to reflect some leniency for lumpy investments. In the inner iteration, the algorithm loops through the set of all generators, both currently built and unbuilt, and modifies the resource mix either by removing a built generator if it makes a negative profit ($< -bound$), or by adding an unbuilt generator k if the addition makes both the generator k and the generators k' of the same technology g make a positive profit ($> bound$ for k'). Note that the termination condition $j = i$ can lead to a solution n_k^{RA}, $\forall k$ which is not a true equilibrium: some generators can have a high profit ($\pi_k^{LT} > bound$), since the only conditions checked in the inner for iteration are $\pi_k^{LT} > 0$ and $\pi_{k'}^{LT} > -bound$; this happens due to the presence of non-convexities.

In the heuristic approach illustrated, different pricing schemes are applied in the operational stage, where investment decisions are considered as fixed. The trial-and-error approach leads to a solution that may be far from the cost-minimizing solution and may vary significantly depending on the pricing scheme chosen and on the order of generators. Furthermore, it does not consider the existing resource mix (greenfield assumption) or

Algorithm 5: Heuristic approach

Data: An instance of the clearing problem plus all data for the GEP
Result: n_k^{RA}, for each k for *e.g.* (2.4)
    // revenue-adequate resource mix for the chosen pricing scheme

begin
  $n_k^*, \forall k \leftarrow$ `Solve` (2.7)   // find the maximum-welfare resource mix
  `Set` $i = 0$; `Set` *bound*
  `Set` $n_k = n_k^*, \forall k$
  `Order` K // ensure technology neutrality
  while $|\pi_k| > bound\ \forall k$ **do**
    `Set` $check(g) = TRUE, \forall g$
    `Set` $j = i$
    for $k \in K$ **do**
      `Set` $g = g_k$
      if $n_k = 0 \wedge check(g) = TRUE$ **then**
        `Set` $n_k = 1$ // try to modify the resource mix
        $(\pi_k^{LT},\ \pi_{k'}^{LT}\ \forall k' \in g) \leftarrow$ `Solve` (2.1), `Solve` *e.g.* (2.4)   // find the long-term profits for all generators of the technology g
        if $\pi_k^{LT} > 0 \wedge \pi_{k'}^{LT} > -bound\ \forall k' \in g$ **then**
          $i \leftarrow i + 1$ // the resource mix has been modified
        else
        `Set` $n_k = 0$ // do not modify the resource mix
        $check(g) = FALSE$
      else if $n_k = 1$ **then**
        $\pi_k^{LT} \leftarrow$ `Solve` (2.1), `Solve` *e.g.* (2.4)   // find the l-t profit for k
        if $\pi_k^{LT} < -bound$ **then**
          `Set` $n_k = 0$ // modify the resource mix
          $i \leftarrow i + 1$
      else   // $n_k^i = 0 \wedge check(g) = FALSE$
    if $j = i$ **then**
      `Set` $n_k^{RA} = n_k \forall k$
      STOP   // Alg terminate because no generator added or removed

investment decisions for renewables. Instead, it only considers scenarios with different penetration levels.

The *Optimization approach* (3), along with investment decisions, endogenously determines market clearing prices to express the Revenue Adequacy (RA) constraint

$$revenues \geq costs$$

Since the optimization model necessitates the endogenization of prices along with the dispatch order to express revenues, the incorporation of any two-step pricing scheme is precluded. As shown in column 1 of Tab. 2.4, the only PRICING models that exhibit independence from the UCED solution (in the price determination) are CH (2.6), LD (2.13), and P-D (2.26), but only the latter determines both the price and the dispatch in a single model. Therefore, the P-D has been extended by [Guo et al., 2022] to be applicable in the long term. As illustrated in Fig. 2.10, a mixed-integer bilinear optimization model is formulated starting from a traditional MIP GEP: the investment decisions n are treated as parameters, while the binary variables γ are relaxed to be continuous in order to derive the dual of the Unit Commitment (UC) relaxed problem; then the integrality conditions are restored, the investment, UC and dual constraints are combined, and the investment and operating costs, together with the weighted duality gap f^{DG} (given by the difference between the primal UC objective and the dual objective) are minimized in the objective function. The hourly clearing prices are the dual variables λ of the balance constraints

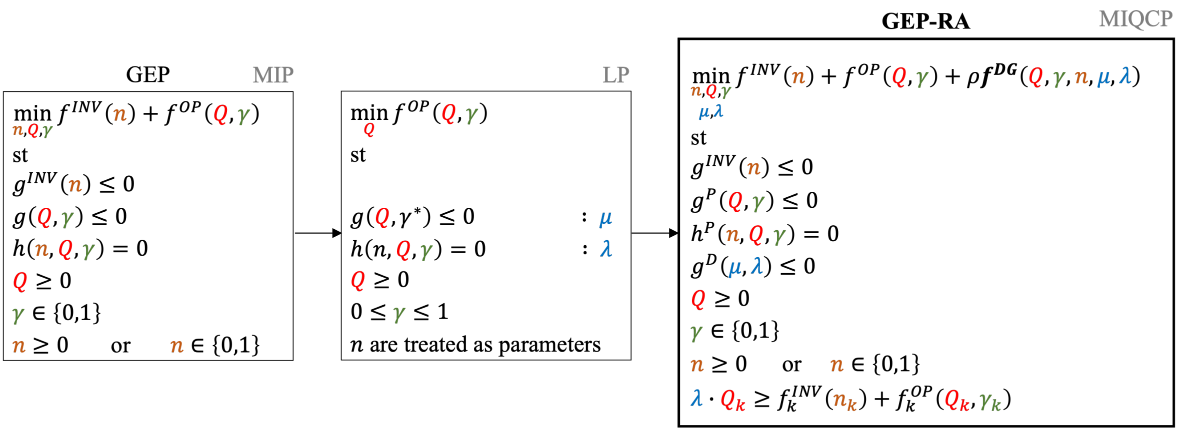


Figure 2.10: Derivation of the GEP-RA model (in the bold block): the first two blocks are only needed to formulate the final model, which is a combination of investment constraints, primal and dual operational constraints, and revenue adequacy constraints

considered by the CP; therefore, they may differ in general from those that the *Market Operator* would determine as they could be pushed up in order to meet the RA constraint.

Since the RA constraint is fulfilled by price spikes, the model loses the explainability of prices (3); moreover, with fixed investment decisions, in the real-time model (which does not take into account the RA constraint) the same spikes will likely not occur, jeopardizing generators' profitability.

In addition to the aforementioned models, [Dvorkin et al., 2017] has proposed an optimization approach that specifically addresses revenue adequacy concerns for electrochemical Energy Storage (ES). The paper presents a bilevel model designed to determine the optimal siting and sizing of ES while ensuring their revenue adequacy. In the lower level, the prices are determined as the dual variables of nodal balance constraints, taking into account dispatch constraints (including ramps) for thermal generators, wind spillage constraints based on wind production forecast, grid constraints, and ES constraints. In the upper level, investment and RA constraints for ES are considered, together with a fixed thermal fleet (no thermal expansion), on which operational constraints limit the values of the binary variables. The linearity of the lower level, resulting from the treatment of binary variables as parameters decided by the upper level, allows its reformulation through the strong duality equality included in the upper level. The division of constraints between upper and lower levels, depending on whether they constrain binaries or not, together with the inclusion of, for example, ramp constraints in the LL, can lead to a price that is difficult to interpret, thereby compromising again the simplicity requirement (3). Furthermore, ensuring revenue adequacy only for ES and its impact on the clearing price may lead to significant losses for thermal or renewable generators. Most importantly, the evolution of batteries should be evaluated in relation to the evolution of renewables, whereas the current model only considers a fixed fleet of wind generators.

Chapter 3

Bilevel optimization models

3.1 Mathematical Programming and optimization taxonomy

Mathematical Programming (MP) serves as a formal language designed to articulate optimization problems and provides a structured framework for their resolution. Within this framework, each formal MP sentence, referred to as a formulation, is constructed from a set of fundamental entities. These entities include index sets and parameters, encapsulating the problem input; decision variables, representing the solution for a particular instance; an objective function, assessing variable assignments; and diverse constraints, defining the feasible decisions within the problem space. The assembly of these components forms the foundation for expressing and solving optimization problems within the MP framework.

Assuming, w.l.o.g, minimization, the MP formulation of a *single-level* generic deterministic optimization problem[1] can be denoted as in (3.1)

$$\min_{x} \quad f(x) \tag{3.1a}$$

$$\text{s.t.} \quad g(x) \leq 0 \tag{3.1b}$$

$$h(x) = 0 \tag{3.1c}$$

$$x \in \mathcal{X} \tag{3.1d}$$

In problem (3.1) a **single decision-maker** decides on x, which represents the vector of variables, according to the objective function $f(x)$ and the inequality (3.1b) and possibly equality (3.1c) set of constraints. Throughout this thesis we will assume that we are

[1] We will concentrate on deterministic problems; there is also a wide class of optimization problems under uncertainty, that can be tackled with Stochastic, Robust, or Chance Constraints, e.g. [Roald et al., 2023], [Van Ackooij et al., 2018] for comprehensive approaches in the energy field.

always optimizing over closed sets. The taxonomy, and related solution complexity, of a specific optimization problem is determined by the type of the real-valued functions, that we assume smooth, present in problem (3.1), i.e. f, g, h, and by the set $\mathcal{X}$. Depending on each one of them, problem (3.1) can be differently categorized and tackled.

In MP, the utilization of reformulation techniques is a common practice. A reformulation $\mathcal{Q}$ of a given problem $\mathcal{P}$ is essentially an auxiliary problem that shares certain properties with $\mathcal{P}$. In the realm of MP, reformulations are typically employed to transform the original problem into a more tractable form, facilitating a more straightforward solution. The reformulation $\mathcal{Q}$ of a problem $\mathcal{P}$ can take various forms to enhance its solvability and better suit specific objectives. These reformulations may:

A **Preserve Optimality Properties**: $\mathcal{Q}$ has the capability to retain all or some of the optimality properties, either at a local or global level, inherent in $\mathcal{P}$;

B **Result from Constraint Removal**: $\mathcal{Q}$ can be derived from $\mathcal{P}$ by strategically eliminating certain constraints, a process commonly known as *relaxations*. This approach aims to simplify the problem structure while maintaining its essential characteristics;

C **Approximate Components of $\mathcal{P}$**: $\mathcal{Q}$ may involve approximations of specific components of $\mathcal{P}$. These approximations, while not guaranteeing optimality, serve the purpose of providing a practical and feasible solution approach. It's noteworthy that such approximations may or may not come with explicit optimality guarantees.

A more comprehensive exploration of reformulations can be found in e.g. [Liberti, 2009] and [Vielma, 2015].

On their side, optimization problems of the form (3.1) can be usually divided[2] into linear, quadratic, generally nonlinear, and non-convex. They can have only continuous variables, only integer variables, or a mix of continuous and integer variables. The most general class is the one of the Mixed-integer Nonlinear Programming (MINLP), e.g. [Sahinidis, 2019]. MINLP, in turn, can be divided into convex MINLP problems or non-convex MINLP problems, depending on the convexity of the related relaxations, e.g. [Burer and Letchford, 2012]. We refer to the simple diagram in Fig. 3.1 for a quite comprehensive taxonomy scheme. Notably for convex formulation and reformulations of

[2]Given that a comprehensive examination of these topics exceeds the scope of this thesis, we will refrain from presenting proofs and instead furnish bibliographic references for those interested in delving deeper into these subjects

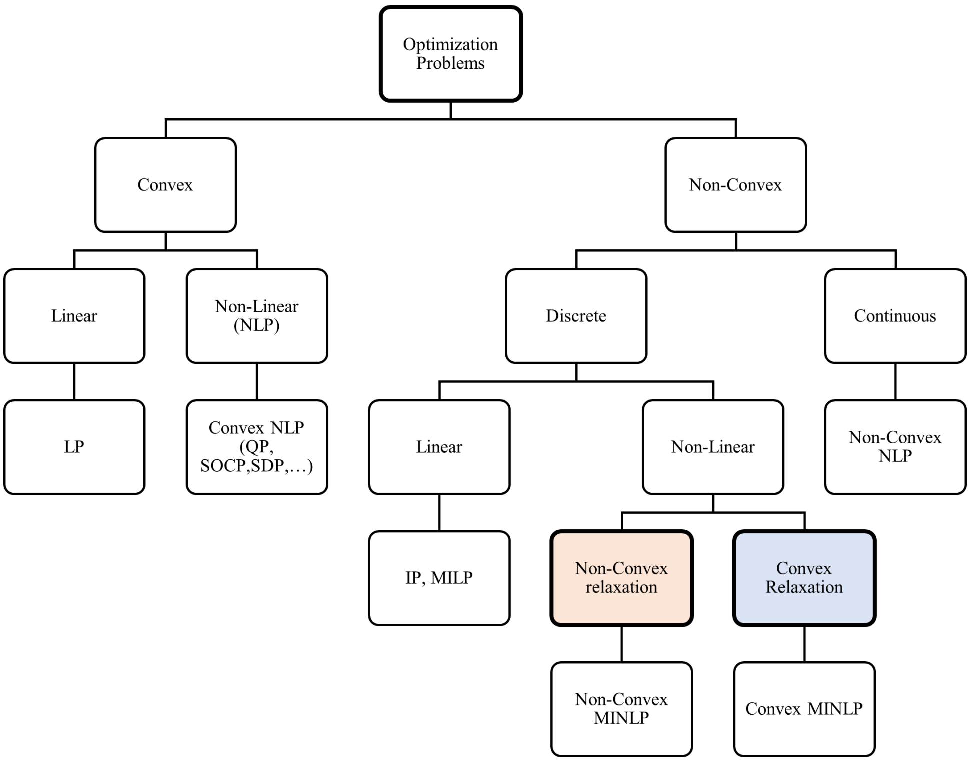


Figure 3.1: Optimization Problem taxonomy

general nonlinear problems we refer to [MOSEK ApS, 2023] for an extensive treatment of relevant sub classes of non-linear convex models such as Second Order Cone Programming (SOCP) and Semi-Definite Programming (SDP).

Among the general non-convex ones are the *bilevel problems*, in which the optimization problem is split into two levels and the decisions are coupled between a so-called upper-level problem and a so-called lower-level. In this thesis, we will basically deal with these classes of optimization problems, therefore it is useful to survey some classes of them. We refer to the recent surveys [Beck and Schmidt, 2021] and [Kleinert et al., 2021a] for a more complete treatment of this important class of optimization models and to [Dempe, 2020] and reference therein also for an extensive list of works in the bilevel optimization.

3.2 Bilevel optimization brief history and generalities

The history of bilevel optimization traces back to Von Stackelberg's leader-follower games in the 1930s [Von Stackelberg, 1934] and later, [Stackelberg, 1952]. Notably, Bracken and McGill [Bracken and McGill, 1973] applied this concept in 1973 to minimize military costs. In 1977, Candler and Norton [Candler and Norton, 1977] recognized the intricate nature of bilevel problems, noting non-convexity and disconnected feasible sets even in cases with all linear components. A fact that at the present time is well understood. Later years saw the formalization of the NP-hardness of linear bilevel problems. Early algorithms, such as Candler and Norton's [Candler and Norton, 1977] enumerative method and Bialas and Karwan's kth-best algorithm [Bialas and Karwan, 1978], emerged in the late 1970s. A pivotal shift occurred in 1981 when Fortuny-Amat and McCarl [Fortuny-Amat and McCarl, 1981] proposed a groundbreaking approach for convex-quadratic bilevel problems. By replacing the follower problem with its Karush–Kuhn–Tucker (KKTs) conditions, due to Harold W. Kuhn and Albert W. Tucker appeared in the seminal paper [Khun and Tucker, 1951][3], the authors transformed it into a single-level problem at least in principle now solvable by standard mixed-integer solvers. This historical narrative underscores the persistent challenges of bilevel optimization and the evolution of algorithms, with early attempts laying the foundation for more effective solutions in subsequent years.

In 2009, a significant milestone in the field of computational bilevel optimization was achieved with the introduction of a branch-and-cut approach specifically designed for purely integer bilevel problems [DeNegre and Ralphs, 2009]. Following this breakthrough, the ensuing decade witnessed a surge in computationally oriented research, addressing various classes of bilevel problems. The introduction of the branch-and-cut approach not only provided a tangible solution for a specific subset of problems but also catalyzed a broader interest in developing efficient computational methods for diverse challenges within the realm of bilevel optimization.

Stepping into some formal detail, a *bilevel* or *hierarchical* optimization problem is an optimization problem constrained by the set of the optimal solutions of another optimization problem. It can be generically denoted as in (3.2)

$$\min_{x\in X, y} \quad F(x,y) \tag{3.2a}$$

$$\text{s.t.} \quad G(x,y) \geq 0 \tag{3.2b}$$

$$y \in S(x) \tag{3.2c}$$

[3] We refer the interested reader also to [Cottle, 2012] for a historical clarification of the role of William Karush with his master's thesis [Karush, 1939]

where $S(x)$ is the set of the optimal solutions of the x-parameterized problem (3.3):

$$\min_{y \in Y} \quad f(x, y) \tag{3.3a}$$

$$\text{s.t.} \quad g(x, y) \geq 0 \tag{3.3b}$$

Therefore, in essence, problems (3.2) and (3.3) are fused into a conceptually unique problem. More in detail, in bilevel optimization problems, such as (3.2)-(3.3), there are **two decision-makers**, each with their own set of variables and constraints. Problem (3.2) is known as the *upper-level* or *leader's problem*, while problem (3.3) is the *lower-level* or *follower's problem*, reflecting a hierarchical structure: any decision (x) taken by the leader to optimize his goals is affected by the response (y) of the follower, who will seek to optimize his own outcomes.

The variables $x \in \mathbb{R}^{n_x}$ are the upper-level variables (or leader's decisions) and $y \in \mathbb{R}^{n_y}$ are lower-level variables (or follower's decisions). The objective functions read $F, f : x \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}$, while the constraints are in the upper-level $G : x \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^m$ and in the lower-level $g : x \in \mathbb{R}^{n_x} \times \mathbb{R}^l \to \mathbb{R}^m$. All functions F, f, G and g are here assumed to be continuously differentiable. The set $X \subseteq \mathbb{R}^{n_x}$ and $Y \subseteq \mathbb{R}^{n_y}$ possibly indicate integrality restrictions of the related variables. As an example if $X = \mathbb{Z}^{n_x}$, then the upper-level problem is an integer program. Lastly, all upper-level variables that are present in the lower-level constraints are usually called linking variables. We give the following useful definitions: 3.2.1, 3.2.2, and 3.2.3.

Definition 3.2.1 *(Shared Constraint Set) The set*

$$\Omega = \{(x, y) \in (X \times Y) : G(x, y) \geq 0,\ g(x, y) \geq 0\} \tag{3.4}$$

is named the shared constraint set.

Definition 3.2.2 *(Bilevel Feasible Set; Inducible Region) The set*

$$\mathcal{F} = \{(x, y) : (x, y) \in \Omega,\ y \in S(x)\} \tag{3.5}$$

is called the bilevel feasible set or inducible region.

Definition 3.2.3 *(High-Point Relaxation) The problem of minimizing the upper-level* *objective function over the shared constraint set, i.e.*

$$\min_{x,y} \quad F(x,y) \tag{3.6a}$$

$$s.t. \quad (x,y) \in \Omega \tag{3.6b}$$

is called the high-point relaxation (HPR) of Problem (3.2)*.*

We can note that the HPR, defined by problem (3.6), is identical to the original bilevel problem (3.2) except for the set valued constraint $y \in S(x)$, i.e., except for the lower-level optimality conditions. In this respect, it is a relaxation of (3.2), and notably a special one. The upper-level decision-maker typically has complete knowledge of the lower-level problem. In contrast, the lower-level decision-maker observes the leader's decisions (the upper-level variables are fixed parameters in the lower-level problem) and optimizes his own strategies. Assuming the leader lacks knowledge of the follower's model, he can learn over time by observing the follower's reactions and thereby gaining insight into the follower's model. This process is commonly referred to as *Learning game* or *Inverse optimization*, see [Borrero et al., 2022]. In case of multiple lower-level optimal solutions y for any given upper-level decision vector x (the set $S(x)$ is not a singleton), the leader can assume an *optimistic position*, expecting the follower to choose that solution from the optimal set which leads to the best upper-level objective function value (the lower-level solution minimizes the upper-level function), or a *pessimistic position*, expecting the follower to choose that solution from the optimal set which leads to the worst upper-level objective function value (the lower-level solution maximizes the upper-level function). The *minimization* written as (3.2a) assumes the optimistic position as the leader controls y, as we will assume in our work. The problem assuming the pessimistic position is in general more difficult to define and can result in the non-existence of a solution. In [Dempe et al., 2018], and [Sinha et al., 2017] the authors review bilevel optimizations and provide additional details on bilevel pessimistic problems.

In this thesis, we will consider linear lower-level problems, that we can write as in (3.7).

$$\min_{y} \quad d^T y \tag{3.7a}$$

$$\text{s.t.} \quad Cx + Dy \geq b \tag{3.7b}$$

$$y \geq 0 \tag{3.7c}$$

In (3.7) $d \in \mathbb{R}^{n_y}$, $C \in \mathbb{R}^{l \times n_x}$, $D \in \mathbb{R}^{l \times n_y}$, and $b \in \mathbb{R}^l$. We note that for the lower-level problem, x is not a variable but a parameter. The dual of the Problem (3.7) is the x-parameterized linear problem (3.8)

$$\max_{\lambda} \quad (b - Cx)^T \lambda \tag{3.8a}$$

$$\text{s.t.} \quad D^T \lambda \leq d \tag{3.8b}$$

$$\lambda \geq 0 \tag{3.8c}$$

3.3 Bilevel optimization single level reformulations

When the lower-level problem is linear, it is possible to replace it with its necessary and sufficient optimality conditions, the KKTs.

Note that for nonlinear lower-level problems, the KKTs are not sufficient optimality conditions; for a discussion of nonlinear lower-level problems, we refer the reader to [Kleinert et al., 2021a]. The KKTs conditions appear as primal, dual, and complementarity constraints, and reduce the overall bilevel optimization problem to a special single-level constrained optimization problem (3.9), named in the literature *Mathematical Program with Complementarity Conditions* (MPCC)

$$\min_{x,y,\lambda} \quad F(x, y) \tag{3.9a}$$

$$\text{s.t.} \quad G(x, y) \geq 0 \tag{3.9b}$$

$$0 \leq (Cx + Dy - b) \perp \lambda \geq 0 \tag{3.9c}$$

$$0 \leq (d - D^T \lambda) \perp y \geq 0 \tag{3.9d}$$

The constrains (3.9c), (3.9d) are equivalent to:

$$\lambda \geq 0 \tag{3.9e}$$

$$Cx + Dy - b \geq 0 \tag{3.9f}$$

$$d - D^T \lambda \geq 0 \tag{3.9g}$$

$$y \geq 0 \tag{3.9h}$$

$$(Cx + Dy - b)^T \lambda = 0 \tag{3.9i}$$

$$(d - D^T \lambda)^T y = 0 \tag{3.9j}$$

By optimizing over x, y, and λ simultaneously in problem (3.9), every global solution of (3.9) corresponds to an optimistic bilevel solution. The constraints (3.9i), (3.9j) are the complementarity slackness conditions: either $Cx + Dy - b = 0$ and $\lambda \geq 0$ or the other way around, $\lambda = 0$ and $Cx + Dy - b \geq 0$. The case $Cx + Dy - b = 0$ and $\lambda = 0$ is also feasible, but generally irrelevant. In the simplest case with F, G, and H linear, the MPCC Problem (3.9) is linear except for the complementarity conditions, which make it a single-level non-convex optimization problem that is $\mathcal{NP}$-hard to solve, [Deng, 1998]. One can eliminate the non-convex terms $x\lambda$ and $y\lambda$, and then solve the model as a mixed-integer linear program (MILP) using different approaches as follows.

An early approach proposed by *Fortuny-Amat and McCarl* [Fortuny-Amat and McCarl, 1981] reformulates the disjunctive complementarity constraints (3.9i), (3.9j) by introducing binary variables $z1$, $z2$ and a well-known approach usually termed Big-M as in (3.10).

$$Cx + Dy - b \leq M\ (1 - z1) \tag{3.10a}$$

$$\lambda \leq M\ z1 \tag{3.10b}$$

$$d - D^T\lambda \leq M\ (1 - z2) \tag{3.10c}$$

$$y \leq M\ z2 \tag{3.10d}$$

$$z1, z2 \in \{0, 1\} \tag{3.10e}$$

If $z1 = 0$, then the dual variable $\lambda = 0$, and the primal inequality $Cx + Dy - b \leq M$ is inactive; if $z1 = 1$, then the primal inequality is active $Cx + Dy - b = 0$ and the dual variable λ can take positive values. Similarly for $z2$. M is a sufficiently large constant (the *big-M*) that is difficult to set because the dual feasible set corresponding to bounded primal feasible sets is unbounded. Moreover, a not sufficiently large constant may cut off some optimal solutions while an exceedingly large constant may produce a weak LP-relaxation, e.g. [Kleinert et al., 2020]. Another equivalent MILP formulation consists in reformulating the complementarity conditions (3.9i), (3.9j) using Special Order Set of type 1 (SOS1), e.g. [Beale and Tomlin, 1970], [Siddiqui and Gabriel, 2013], i.e., a set of variables where at most one variable can have a value greater than zero as in (3.11).

$$s_1 = Cx + Dy - b \tag{3.11a}$$

$$s_2 = \lambda \tag{3.11b}$$

$$t_1 = d - D^T\lambda \tag{3.11c}$$

$$t_2 = y \tag{3.11d}$$

$$\{s_1, s_2\} \in SOS1 \tag{3.11e}$$

$$\{t_1, t_2\} \in SOS1 \tag{3.11f}$$

The pairs $\{s_1, s_2\}$ and $\{t_1, t_2\}$ are declared as SOS1: of each pair, only one variable can take positive values.

A possible alternative to reformulate the MPCC involves exploiting *strong duality* for the follower's linear problem to convert the problem into a single-level quadratic program that is non-convex as in (3.12).

$$\min_{x,y,\lambda} \quad F(x, y) \tag{3.12a}$$

$$\text{s.t.} \quad G(x, y) \geq 0 \tag{3.12b}$$

$$Cx + Dy - b \geq 0 \tag{3.12c}$$

$$\lambda \geq 0 \tag{3.12d}$$

$$d - D^T\lambda \geq 0 \tag{3.12e}$$

$$y \geq 0 \tag{3.12f}$$

$$y^T d = (b - Cx)^T\lambda \tag{3.12g}$$

In Problem (3.12), the complementarity conditions (3.9i), (3.9j) are replaced by the strong-duality equality (3.12g), requiring the primal objective function to be equal to the dual objective function. To demonstrate the equivalence of the two approaches, one can substitute the expression derived from constraint (3.9j)

$$(d - D^T\lambda)^T y = 0 \Longleftrightarrow d^T y = \lambda^T Dy \Longleftrightarrow y^T d = y^T D^T\lambda$$

into the constraint (3.9i)

$$(Cx + Dy - b)^T\lambda = 0 \Longleftrightarrow y^T D^T\lambda = (b - Cx)^T\lambda \Longleftrightarrow y^T d = (b - Cx)^T\lambda$$

Similar to (3.9), the problem (3.12) is non-convex due to the bilinear product of variables $x^T\lambda$. However, the introduction of binary auxiliary variables in this case does not allow an equivalent MILP formulation to be derived, but only an approximate formulation derived from the binary expansion of one of the continuous variables of the bilinear product, [Dias Garcia et al., 2023], [Pereira et al., 2005], [Zare et al., 2019].

Given the complexity of solving bilevel problems exactly, heuristic solution methods have also been developed. These methods do not provably compute a global minimum in finite time. For example, the *Regularization approach* reformulates the complementarity conditions (3.9i), (3.9j) by introducing small scalars $r1, r2$ and solving the resulting nonlinear optimization problem iteratively:

$$(Cx + Dy - b)^T\lambda \leq r1$$
$$(d - D^T\lambda)^T y \leq r2$$

At each iteration, the locally optimal solution is used as the warm start point for the next iteration while the scalar values are reduced. Another example of heuristic approaches is the *Penalty Alternating Direction Method*, which relaxes the strong duality (3.12g) by penalizing it in the objective function and then splits the problem into two subproblems, one for x and one for λ. See [Bylling, 2018] and also [Beck and Schmidt, 2021] for an in-depth explanation of these heuristic methods.

Chapter 4

Development of bilevel models for GEP

In this chapter, we describe the process undertaken to develop the proposed models. In Section (4.1), we delve into the intricacies of this development process, outlining various modeling assumptions. Subsequently, in Section (4.2), we introduce the two bilevel models designed to determine a revenue-adequate configuration for the system, each providing distinct perspectives on optimization objectives.

4.1 Stepwise evolution of the models

As explained in Section (2.1), a meaningful assessment of revenue-adequate investments requires the consideration of appropriate pricing. In order to define the best resource mix for Italy, we first examined the market clearing model used by the (Nominated) Italian Market Operator (GME), as outlined in Section (4.1.1), and then introduced necessary modifications. In particular, Section (4.1.2) introduces a refined version of the market clearing model, that specifically addresses the minimum output of thermal generators. This adjustment will allow startup costs to be taken into account in ensuring revenue adequacy. Additionally, Section (4.1.4) delves into renewable cannibalization and outlines the evolution of the renewable offer strategies; the hypothesized evolution requires the consideration of partially accepted offers. Section (4.1.5) outlines modeling assumptions for the integration of batteries into the system to store excess renewable production and use it in later hours. Section (4.1.6) emphasizes the importance of designing a system that is fully adequate and accommodates possible side payments. Finally, Section (4.1.7) illustrates the cases in which the model under consideration results in price indeterminacy, along with modeling adjustments aimed to better reflect real-world investment decisions.

4.1.1 The Italian Market Clearing problem

The Italian Market Operator receives a set S of selling offers from potential sellers and a set B of buying bids from potential buyers, each in the form of a pair, and referred to a specific hour. Selling offers are pairs $(\overline{SQ}_j, SP_j)$ indicating that the seller is willing to sell up to $\overline{SQ}_j$ units of electricity at unit price SP_j, while buying bids are pairs $(\overline{BQ}_i, BP_i)$ indicating that the buyer is willing to buy up to $\overline{BQ}_i$ units of electricity at unit price BP_i. In order to preserve the convexity of the model, if one seller issues two offers $(\overline{SQ_1}, SP_1)$ and $(\overline{SQ_2}, SP_2)$ with $SP_1 < SP_2$, then he is willing to sell up to $\overline{SQ_1}$ units of electricity at unit price SP_1, and, if all those are sold, he is also willing to sell up to $\overline{SQ_2}$ units at the larger unit price SP_2. The submitted offers/bids on the market compose two curves: the *supply curve*, increasing with respect to the offer price SP_j, for the selling offers, and the *demand curve*, decreasing with respect to the bid price BP_i, for the buying bids, as illustrated in Fig. 4.1.

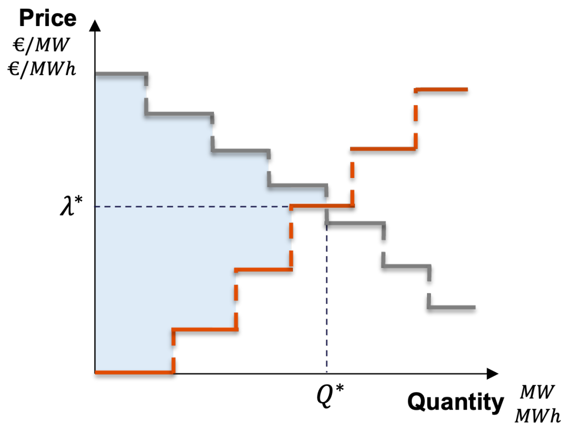


Figure 4.1: Supply (orange) and demand (gray) curves: at the intersection, λ^* represents the price of electricity and Q^* the total quantity accepted; the area (blue) represents the social welfare

In non-pathological situations, the curves meet at a single point that defines

- the **electricity price** λ^*, corresponding to the price of the *marginal*, i.e. the most costly selling offer accepted, and
- the total quantity Q^* of purchased energy.

The price λ^* is paid to *all* the sellers whose offers contribute to the total accepted quantity Q^* (*pay-as-cleared mechanism*).

From a mathematical point of view, the MO clears the day-ahead market by solving the primal problem **MGP with Elastic demand** (MGP-E) (4.1) for each hour to which the bids are referred. In the model, a system with one zone is considered as if the transmission system had no bottlenecks, i.e. the so-called single bus system. The model with multiple zones as envisioned by the Italian regulation will be presented shortly.

$$\max_{BQ_i \geq 0, SQ_j \geq 0} \sum_{i \in B} BP_i BQ_i - \sum_{j \in S} SP_j SQ_j \tag{4.1a}$$

$$BQ_i \leq \overline{BQ}_i \qquad i \in B \qquad (\eta_i \geq 0) \tag{4.1b}$$

$$SQ_j \leq \overline{SQ}_j \qquad j \in S \qquad (\mu_j \geq 0) \tag{4.1c}$$

$$\sum_{i \in B} BQ_i = \sum_{j \in S} SQ_j \qquad (\lambda) \tag{4.1d}$$

In (4.1), the optimal values BQ_i^* and SQ_j^* represent respectively how much of the i-th buying bid/j-th selling offer is accepted by the MO. Constraints (4.1b) and (4.1c) limit the accepted quantity with respect to the bid quantity while (4.1d) requires that whatever is sold is also bought (*supply-demand balance constraint*). The objective function (4.1a) conceptually maximizes *social welfare*, geometrically defined as the area between the demand and the supply curves.

Let η_i and μ_j be the dual variables associated with the constraints (4.1b) and (4.1c) respectively, while λ is the dual variable associated with the constraint (4.1d). The dual of (4.1) is the problem (4.2).

$$\min_{\eta_i \geq 0, \mu_j \geq 0, \lambda} \sum_{i \in B} \overline{BQ}_i \eta_i + \sum_{j \in S} \overline{SQ}_j \mu_j \tag{4.2a}$$

$$\eta_i + \lambda \geq BP_i \qquad i \in B \qquad BQ_i \geq 0 \tag{4.2b}$$

$$\mu_j - \lambda \geq -SP_j \qquad j \in S \qquad SQ_j \geq 0 \tag{4.2c}$$

The Complementarity Slackness Conditions (CSC) of (4.1) and (4.2) read as in (4.3).

$$\eta_i(\overline{BQ}_i - BQ_i) = 0 \qquad i \in B \qquad \mu_j(\overline{SQ}_j - SQ_j) = 0 \qquad j \in S \tag{4.3a}$$

$$BQ_i(\eta_i + \lambda - BP_i) = 0 \qquad i \in B \qquad SQ_j(\mu_j - \lambda + SP_j) = 0 \qquad j \in S \tag{4.3b}$$

They define the following rejection or acceptance criteria:

Bid	rejected	$BQ_i^* = 0$	$\lambda^* \geq BP_i - \eta_i^*$	$\eta_i^* = 0$	$\lambda^* \geq BP_i$
	part. accepted	$0 < BQ_i^* < \overline{BQ}_i$	$\lambda^* = BP_i - \eta_i^*$	$\eta_i^* = 0$	$\lambda^* = BP_i$
	fully accepted	$BQ_i^* = \overline{BQ}_i$	$\lambda^* = BP_i - \eta_i^*$	$\eta_i^* \geq 0$	$\lambda^* \leq BP_i$
Offer	rejected	$SQ_j^* = 0$	$\lambda^* \leq SP_j + \mu_j^*$	$\mu_j^* = 0$	$\lambda^* \leq SP_j$
	part. accepted	$0 < SQ_j^* < \overline{SQ}_j$	$\lambda^* = SP_j + \mu_j^*$	$\mu_j^* = 0$	$\lambda^* = SP_j$
	fully accepted	$SQ_j^* = \overline{SQ}_j$	$\lambda^* = SP_j + \mu_j^*$	$\mu_j^* \geq 0$	$\lambda^* \geq SP_j$

In particular:

- from primal and dual feasibility and (4.3a), one has that whenever $\lambda^* < BP_i$ we have $\eta_i^* > 0$, i.e., the market clearing price is smaller than the price at which the buying bid i would be accepted, then $BQ_i^* = \overline{BQ}_i$, i.e., the buying bid is fully accepted.
- Dually, whenever $\lambda^* > SP_j$ it must result $\mu_j^* > 0$, i.e., the market clearing price is larger than the price at which selling offer j would be accepted, then $SQ_j^* = \overline{SQ}_j$, i.e., the selling offer is fully accepted.

So, the objective function (4.2a) measures – and minimizes – the deviation between the amount of money that the buyers were willing to pay in order to get the amount of electricity they actually received, plus the amount of money that the buyers were expecting to get in order to sell the amount of electricity they actually sold, and that corresponding to forcing everybody to sell/buy at the fixed market clearing price λ^*. Of course, nobody can, and is forced to, buy/sell at a price that is too high/low for his expectations; in fact, due to (4.3b), a buying bid for which $\lambda^* > BP_i \Rightarrow \eta_i^* = 0 > BP_i - \lambda^*$ results in $BQ_i^* = 0$, and analogously for selling offers.

The dual variable η_i represents the operating profit per unit: it is null when the i-th buying bid is either rejected or partially accepted or non-negative when the bid i is totally accepted:

$$\eta_i^* = \max\{0, BP_i - \lambda^*\}$$

Similarly, μ_j^* represents the **operating profit per unit** of the j-th selling offer:

$$\mu_j^* = \max\{0, \lambda^* - SP_j\}$$

Since $\eta_i^*, \mu_j^* > 0$ only when $BQ_i^* = \overline{BQ}_i$ and $SQ_j^* = \overline{SQ}_j$ respectively, the optimal value of the dual objective function (4.2a) represents the sum of the consumers' profits $\sum_{i \in B} \overline{BQ}_i \eta_i^*$ and the producers' profits $\sum_{j \in S} \overline{SQ}_j \mu_j^*$, that is in turn equal to the social welfare for the strong-duality equality (the primal objective function represents the social welfare).

The value of the optimal market clearing price can result in any of the four cases illustrated in Fig. 4.2, based on the characteristics of the intersection point [GME, 2002]:

E1 *unique price* defined by the marginal supply offer $\lambda^* = SP_{j^*}$

E2 *unique price* defined by the marginal buying bid $\lambda^* = BP_{i^*}$

E3 *quantity-indeterminacy*: $\lambda^* = SP_{j^*} = BP_{i^*}$; any quantity in $[Q^*_{min}, Q^*_{max}]$ is optimal

E4 *price-indeterminacy*: any price in $[\lambda^*_{min}, \lambda^*_{max}]$ is optimal

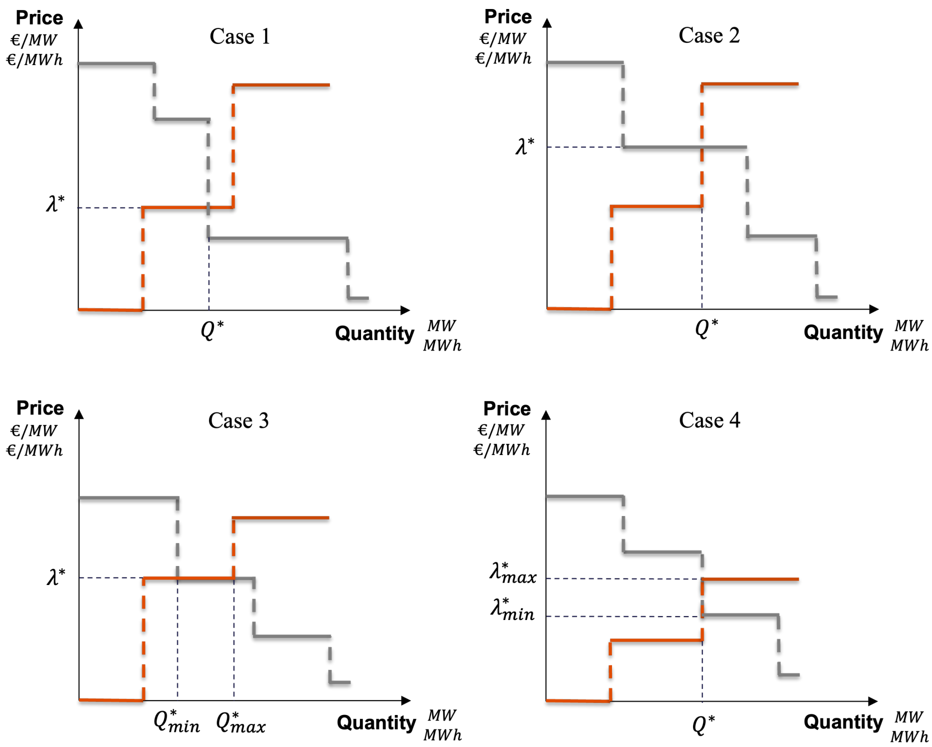


Figure 4.2: Market clearing price and total quantity accepted under different assumptions about the characteristics of the intersection of the supply (orange) and demand (gray) curves

Both cases (**E3**) and (**E4**) are caused by degenerate solutions in the primal and dual problems respectively. In (**E4**), social welfare does not change depending on which price is chosen in $[\lambda^*_{min}, \lambda^*_{max}]$; what changes is the ratio of consumer to producer profits: by choosing λ^*_{min}, consumer profits increase [Byers and Hug, 2023].

Given the nature of the electricity, the demand curve can also be imagined as *rigid*, i.e., independent of the price, and therefore be a simple constant D with a possible cap at the price given by an estimation of the Value of Lost Load (VoLL), which in the Italian market is conventionally set at 3000 €/MWh [European Commission, 2020]. In this case, the number of possible intersection cases reduces to three, as shown in Fig. 4.3:

I1 *unique price* defined by the marginal generator $\lambda^* = P_{k^*}$

I2 *energy-not-provided*: the demand exceeds the total quantity offered, giving rise to Energy Not Provided (ENP), and thus to a market clearing price equal to the VoLL ($\lambda^* = VoLL$)

I3 *price-indeterminacy*: any price in the interval $[\lambda^*_{min}, \lambda^*_{max}]$ is optimal

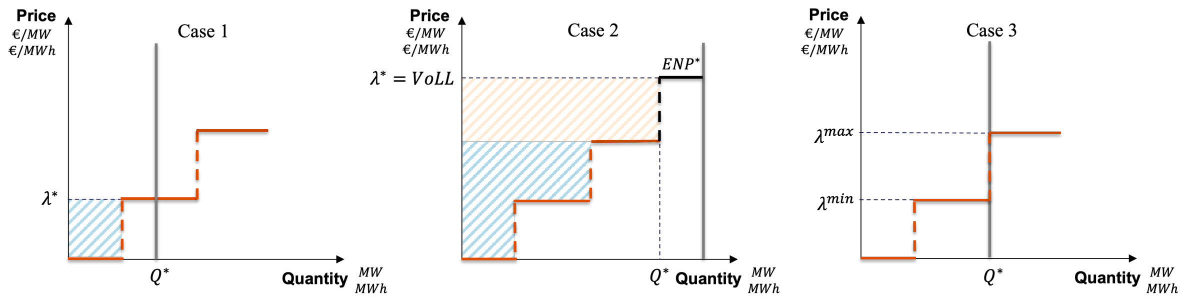


Figure 4.3: Market clearing price and total quantity accepted under different levels of inelastic demand (gray vertical line): in Case 1 there are infra-marginal rents (blue area), in Case 2 there are both infra-marginal (blue area) and scarcity (orange area) rents, while in Case 3 the infra-marginal rents depend on the price chosen from the indeterminacy interval

The MO solves the primal problem **MGP with Inelastic demand** (MGP - I) (4.4).

$$\min_{Q_k \geq 0, ENP \geq 0} \sum_{k \in K} P_k Q_k + VoLL\ ENP \tag{4.4a}$$

$$Q_k \leq \overline{Q}_k \qquad k \in K \qquad (\mu_k \geq 0) \tag{4.4b}$$

$$\sum_{k \in K} Q_k + ENP \ = \ D \qquad (\lambda) \tag{4.4c}$$

The non-negative variable ENP is introduced in the supply-demand balance constraint (4.4c) and it is penalized in the objective function (4.4a) at VoLL.

The dual of (4.4) is the problem (4.5).

$$\max_{\mu_k \geq 0, \lambda} D\lambda - \sum_{k \in K} \overline{Q}_k \mu_k \tag{4.5a}$$

$$\mu_k - \lambda \geq -P_k \qquad k \in K \qquad (Q_k \geq 0) \tag{4.5b}$$

$$\lambda \leq VoLL \qquad (ENP \geq 0) \tag{4.5c}$$

In the objective function (4.5a) $D\lambda$ represents the consumers' expenditure while the term $-\sum_{k \in K} \overline{Q}_k \mu_k$ forces the solution to assign a value $\mu^*_k > 0$ to the bids with the lowest incremental cost. In other words, the cheapest offers needed to cover the demand are fully accepted

($Q_k^* = \overline{Q}_k$). Even with inelastic demand, the dual variable $\mu_k^* = \max\{0, \lambda^* - P_k\}$ represents the **operating profit per unit**: in particular, it is positive only if the generator is *infra-marginal* ($\lambda^* > P_k$) or in the case of scarcity rent created by $ENP^* > 0$ ($\lambda^* = VoLL > P_k$); in all other cases, it is zero.

We now assume that each selling offer (generator) k is associated with a given zone among a set Z of possible ones by the set K_z. Exchanges of electricity within the same zone are not limited; however, a set L of transmission lines between zones is given, and exchanges are limited by the equivalent capacity of these lines. Unlike other European countries (with the exception of Scandinavia), where zones typically correspond to the entire national territory (each Country a zone), Italy considers $|Z| = 7$ distinct market zones [Terna, 2021], linked as shown in Fig. 4.4.

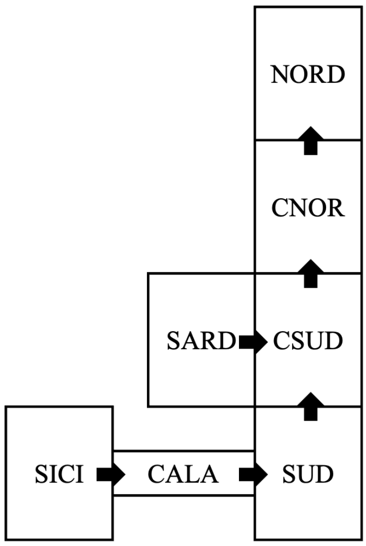


Figure 4.4: Italian market zones and equivalent transmission lines; the depicted direction follows the convention of indicating a positive sign for the flow of energy in that direction and a negative sign for the flow of energy in the opposite direction

This approach aims to account for the specific geographical configuration of the productions and loads and tailor purchase prices based on the balance between generation capacity and demand, which may vary from zone to zone, signaling appropriate pricing differences across zones.

Considering multiple zones, we obtain the primal problem **Constrained MGP with In-**

elastic demand (C-MGP-I) (4.6). For simplicity, we ignore the ENP.

$$\min_{Q_k \geq 0, F_l} \sum_{k \in K} P_k Q_k \tag{4.6a}$$

$$Q_k \leq \overline{Q}_k \qquad k \in K \qquad (\mu_k \geq 0) \tag{4.6b}$$

$$\sum_{k \in K_z} Q_k + \sum_{l \in BS_z} F_l \; = \; D_z + \sum_{l \in FS_z} F_l \qquad z \in Z \qquad (\lambda_z) \tag{4.6c}$$

$$\underline{F}_l \leq F_l \leq \overline{F}_l \qquad l \in L \qquad (\underline{\upsilon}_l, \overline{\upsilon}_l \geq 0) \tag{4.6d}$$

In constraint (4.6d), $\underline{F}_l$ and $\overline{F}_l$ are respectively the minimum and the maximum capacities of the transmission line l (typically, $\underline{F}_l < 0$ and $\overline{F}_l > 0$). The constraints (4.6c) ensure the supply-demand balance at each zone z: the left-hand side represents the energy sources given by generators and incoming flows (arcs entering the zone z), while the right-hand side describes the energy uses given by the zonal demand and outgoing flows (arcs leaving the zone z).

The corresponding dual problem of (4.6) is (4.7).

$$\max_{(\mu_k, \underline{\upsilon}_l, \overline{\upsilon}_l) \geq 0, \lambda} \sum_{z \in Z} D_z \lambda_z - \sum_{k \in K} \overline{Q}_k \mu_k + \sum_{l \in L} (\underline{F}_l \underline{\upsilon}_l - \overline{F}_l \overline{\upsilon}_l) \tag{4.7a}$$

$$\mu_k - \lambda \geq -P_k \qquad k \in K \qquad (Q_k \geq 0) \tag{4.7b}$$

$$\lambda_{z_l^{tail}} - \lambda_{z_l^{head}}) - \underline{\upsilon}_l + \overline{\upsilon}_l = 0 \qquad l \in L \qquad (F_l) \tag{4.7c}$$

The optimal value λ_z^* is the **market clearing price of zone** z. Zonal prices may vary across different zones. Line $l \in L$ is defined by the ordered pair of nodes (z_l^{tail}, z_l^{head}), i.e. a positive flow is from z_l^{tail} to z_l^{head} and a negative flow is from z_l^{head} to z_l^{tail}.

The CSC (4.8) are

$$\underline{\upsilon}_l (F_l - \underline{F}_l) = 0 \qquad l \in L \qquad \overline{\upsilon}_l (\overline{F}_l - F_l) = 0 \qquad l \in L \tag{4.8}$$

and together with the dual constraint (4.7c) define the following criteria on the prices among zones:

$\underline{F}_l < F_l^* < \overline{F}_l$	$\underline{\upsilon}_l^* = 0$	$\overline{\upsilon}_l^* = 0$	$\lambda^*_{z_l^{tail}} = \lambda^*_{z_l^{head}}$
$F_l^* = \underline{F}_l$	$\underline{\upsilon}_l^* = \lambda^*_{z_l^{tail}} - \lambda^*_{z_l^{head}} \geq 0$	$\overline{\upsilon}_l^* = 0$	$\lambda^*_{z_l^{tail}} \geq \lambda^*_{z_l^{head}}$
$F_l^* = \overline{F}_l$	$\underline{\upsilon}_l^* = 0$	$\overline{\upsilon}_l^* = \lambda^*_{z_l^{head}} - \lambda^*_{z_l^{tail}} \geq 0$	$\lambda^*_{z_l^{tail}} \leq \lambda^*_{z_l^{head}}$

As intuition suggests, in the absence of congestion, the price for the producers is unique. Conversely, when congestion materializes in the solution, prices are consequently higher in zones with lower supply (import zones) and vice versa. From a generation standpoint, this also incentivizes investors to develop capacity in the zone where it is needed to reduce congestion (*locational price signals*).

The dual variables $\underline{\upsilon}_l^*$ and $\overline{\upsilon}_l^*$ represent the **congestion rents**:

$$\underline{\upsilon}_l^* = \max\{0, \lambda^*_{z_l^{tail}} - \lambda^*_{z_l^{head}}\} \quad \overline{\upsilon}_l^* = \max\{0, \lambda^*_{z_l^{head}} - \lambda^*_{z_l^{tail}}\} \tag{4.9}$$

The rules of the Italian electrical market require the MO to produce:

- zonal prices λ_z^* for *sellers*, taking into account the extra cost corresponding to the presence of transmission constraints;
- a unique market clearing price λ^* for *buyers*, referred to as PUN (Prezzo Unico Nazionale), ensuring that buyers are neither penalized nor favored based on their geographical location.

4.1.2 Introducing the minimum power output

The Italian MO does not consider technical minimum outputs of thermal generators and thus MGP-I (4.4) may result in generators dispatched for a lower level than what is physically possible. However, after MGP clearing the resulting schedule can be changed by defining new offers on a subsequent market, *the intra-day market* (MI). In so doing generators have the opportunity to adjust their position before the real time, where the schedule defined by the TSO becomes binding. In essence, the producer may decide to:

- re-assign the energy sold in MGP by its own generators (if he has more than one generator in his portfolio);
- submit new offers with other producers to buy back that energy sold in MGP or sell more energy in MI markets;
- pay penalties on the ancillary services market according to the imbalance they produced.

Let's consider the example in Tab. 4.1 with prices in [€/MWh] and quantities in [MWh], assume a rigid system demand of $D = 30$ MWh. The system configuration includes two generators, k_1 and k_2, with offer prices of 5 €/MWh and 10 €/MWh, and offer quantities of 20 MW and 40 MW, respectively; generator k_2 presents a technical minimum level of 15 MW. As illustrated in the left-hand figure of Fig. 4.5, MGP-I dispatches the cheapest generator at capacity and the more expensive generator for the remaining load: $Q_2^* = 10 < \underline{Q}_2 = 15$, therefore its outcome is infeasible due to the technical minimum. Generator k_2 may alternatively:

	k_1	k_2
$\underline{Q}_k$	0	15
$\overline{Q}_k$	20	40
P_k	5	10

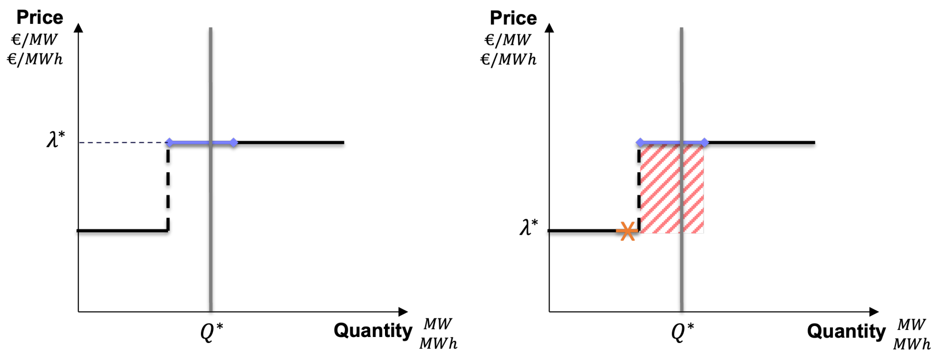
Table 4.1: Economical and physical characteristics of the generators considered in the example

a. on MI try to sell an additional amount of energy $\underline{Q}_2 - Q_2^* = 5$, to reach a feasible point of production, to generator k_1, which would reduce its production to $Q_1{}^* - (\underline{Q}_2 - Q_2{}^*) = 15$ without incurring the cost of production of $Q_1 = 5$ MWh at 5 €/MWh (assuming that the generator offers at its marginal cost);

b. operate at the technical minimum and incur a positive imbalance;

c. choose to stay off, resulting in a negative imbalance.

In the latter two cases, (b.) and (c.), it will pay imbalance prices that will form in the ancillary services markets according to the imbalance rules in force. In case (a.), for generator k_1 to



Q_1^*	Q_2^*	λ^*
20	10	10

γ_1	γ_2	Q_1^*	Q_2^*	λ^*
1	1	15	15	5

Figure 4.5: Comparison between the optimal solutions of MGP-I (left) and MGP+MI (right) with given statuses $\gamma_1 = \gamma_2 = 1$. To meet the technical minimum requirement (violet segment), MGP+MI decreases the dispatch of the cheapest generator (orange segment), which becomes marginal: the reduced price results in a loss (red area) for the more expensive generator.

accept the MI offer, generator k_2 must choose an *offer price* ≤ 5, i.e. below the cost that k_1 would incur if producing instead of buying from k_2. Note that although k_2 is losing money, its opportunity cost is related to choices (b.) or (c.).

Therefore, downstream of MI, the price is $\lambda = 5$, at which k_1 is indifferent between producing the dispatched quantity on MGP and modifying its production.

We approximate the operation of the MGP and MI markets by introducing in the MGP-I model (4.4) a *parameter* γ_k, representing the **commitment status** of the generator k: in the **MGP+MI** model (4.10), $\gamma_k = 1$ indicates that generator k is ON and must be dispatched at a level between the technical minimum $\underline{Q}_k$ and the capacity $\overline{Q}_k$; $\gamma_k = 0$ indicates that generator k is OFF and does not offer.

$$\min_{Q_k \geq 0, ENP \geq 0} \sum_{k \in K} P_k Q_k + VoLL\ ENP \tag{4.10a}$$

$$\underline{Q}_k \gamma_k \leq Q_k \leq \overline{Q}_k \gamma_k \qquad k \in K \qquad (\underline{\mu}_k, \overline{\mu}_k \geq 0) \tag{4.10b}$$

$$\sum_{k \in K} Q_k + ENP \ = \ D \qquad (\lambda) \tag{4.10c}$$

Considering γ_k as a parameter, the dual of (4.10) is the problem (4.11).

$$\max_{(\underline{\mu}_k, \overline{\mu}_k) \geq 0, \lambda} D\lambda + \sum_{k \in K} (\underline{Q}_k \gamma_k \underline{\mu}_k - \overline{Q}_k \gamma_k \overline{\mu}_k) \tag{4.11a}$$

$$-\underline{\mu}_k + \overline{\mu}_k - \lambda \geq -P_k \qquad k \in K \qquad (Q_k \geq 0) \tag{4.11b}$$

$$\lambda \leq VoLL \qquad (ENP \geq 0) \tag{4.11c}$$

The CSC (4.12) are

$$\underline{\mu}_k (Q_k - \underline{Q}_k \gamma_k) = 0 \qquad k \in K \qquad \overline{\mu}_k (\overline{Q}_k \gamma_k - Q_k) = 0 \qquad k \in K \tag{4.12a}$$

$$Q_k(-\underline{\mu}_k + \overline{\mu}_k - \lambda + P_k) = 0 \qquad k \in K \qquad ENP(VoLL - \lambda) = 0 \tag{4.12b}$$

and introduce the following criteria

$\gamma_k = 0$	$Q_k^* = 0$	$\underline{\mu}_k^* \geq 0$	$\overline{\mu}_k^* \geq 0$	$\overline{\mu}_k^* - \underline{\mu}_k^* \geq \lambda^* - P_k$
$\gamma_k = 1$	$Q_k^* = \underline{Q}_k$	$\underline{\mu}_k^* = P_k - \lambda^* \geq 0$	$\overline{\mu}_k^* = 0$	$\lambda^* \leq P_k$
	$\underline{Q}_k < Q_k^* < \overline{Q}_k$	$\underline{\mu}_k^* = 0$	$\overline{\mu}_k^* = 0$	$\lambda^* = P_k$
	$Q_k^* = \overline{Q}_k$	$\underline{\mu}_k^* = 0$	$\overline{\mu}_k^* = \lambda^* - P_k \geq 0$	$\lambda^* \geq P_k$

For $\gamma_k = 1$ the dual variables $\underline{\mu}_k$ and $\overline{\mu}_k$ represent the **loss per unit** and the **profit per unit**, respectively:

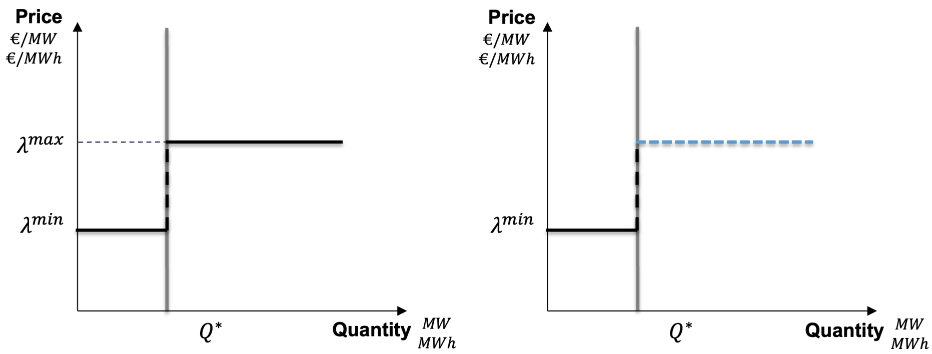
$$\underline{\mu}_k^* = \max\{0, P_k - \lambda^*\} \qquad \overline{\mu}_k^* = \max\{0, \lambda^* - P_k\} \tag{4.13}$$

For $\gamma_k = 0$, on the other hand, the values of $\underline{\mu}_k^*$ and $\overline{\mu}_k^*$ do not impact the objective function

(4.11a) and can therefore take on any value. In this case, they do not represent losses or profits. The only constraint is on the difference $\overline{\mu}_k^* - \underline{\mu}_k^*$.

Assuming $\gamma_1 = \gamma_2 = 1$ in the example, MGP+MI (4.10) leads to the optimal dispatch shown in the right-hand figure of Fig. 4.5: the generator k_1 is no longer dispatched at capacity (reduction of the dispatched quantity of the cheapest generator) while the generator k_2 is dispatched at the technical minimum. The electricity price does not reflect the offer price of k_2 since k_2 is not the marginal generator, although it is the more expensive one dispatched: an additional unit of load would be covered by k_1. Therefore, generator $k1$ bears a loss per unit $\underline{\mu}_2^* = 5$€/MWh.

Another important difference between MGP-I and MGP+MI concerns price indeterminacy cases: in MGP-I, the MO problem sees all generators, even those that are not dispatched, allowing them to cover any additional unit of demand. In contrast, MGP+MI treats off-line generators ($\gamma_k = 0$) as non-existent, making it impossible for them to meet an additional unit of demand and resulting in unmet energy demand and a price at the VoLL. Let's consider the previous example with a demand $D = 20$ MWh: Fig. 4.6 shows prices resulting from MGP-I and MGP+MI.



Q_1^*	Q_2^*	λ^*
20	0	[5,10]

γ_1	γ_2	Q_1^*	Q_2^*	λ^*
1	0	20	0	[5,VoLL]

Figure 4.6: Comparison of the optimal solutions of MGP-I (left) and MGP+MI (right) in the case of price indeterminacy: in MGP-I, the second generator could satisfy one additional unit of demand, whereas in MGP+MI, any additional demand could not be satisfied by the off-line generator ($\gamma_2 = 0$), leading to a price fixed at VoLL

The optimal value of the *binary variable* γ_k, and other variables, are defined in another optimization problem, that considers optimal dispatched quantities and prices from MGP+MI (4.10). The resulting problem is therefore a *bilevel model* in which:

- the lower level (LL) is the problem MGP+MI (4.10) solved for each hour by the MO. This model will be also modified later on;
- the upper level (UL) is the problem of a CP that determines γ_k^*, and other variables,

according to its objective function and constraints. The formulations of the UL will be detailed also subsequently.

Among its constraints, the UL takes into account UC constraints enforcing in each hour t consistency between statuses $\gamma_{k,t}$ and variables representing startup $\alpha_{k,t}$ and shutdown $\beta_{k,t}$ decisions. In this way, it is able to account for startup costs C_k^{SU} in the Revenue Adequacy (RA) constraints, that from the operational point of view can be defined in (4.14) as a first attempt.

$$\sum_t \lambda_t Q_{k,t} \geq \sum_t (C_k^{SU} \alpha_{k,t} + C_k^M Q_{k,t}) \tag{4.14}$$

Constraints (4.14) are needed for each generator $k \in K$, C_k^M is the marginal cost of the generator k.

These RA constraints will later be modified to include investment costs and zonal prices and other components.

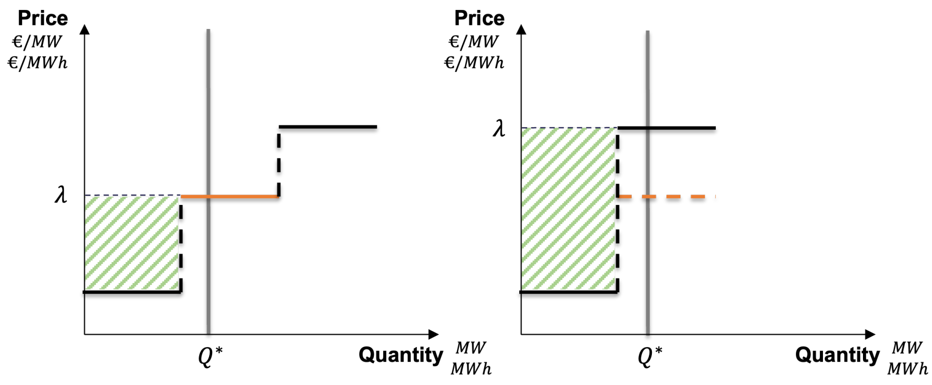
To account for startup/shutdown dynamics (and thus for the startup costs), the technical minimum values $\underline{Q}_k$ must be positive. Without this condition, if the generator is turned ON ($\alpha_{k,t} = 1$), there is no incentive to switch it OFF and pay the startup cost in later hours, as it can operate with zero production ($Q_{k,t} = 0$) even in the ON state ($\gamma_{k,t} = 1$), as per the constraint (4.10b).

Including the technical minimum constraint $Q_{k,t} \geq \underline{Q}_k \gamma_{k,t}$ in the UL instead of the LL would introduce a significant distortion in market outcomes. In fact, the model could result in allocating the entire demand to the more expensive generator while keeping the cheapest OFF.

4.1.3 Strategic withholding

The CP, representing the producers, may decide to adopt a capacity withholding strategy (2) by shutting down some generators (setting $\gamma_k = 0$) to allow generators with higher marginal costs (offer prices) to set the clearing price, and thus satisfy the RA constraints, as illustrated in Fig. 4.7. This is equivalent in reality to the operator owning both the first and the second generator deciding not to offer the capacity of the second generator in order to obtain a higher profit for the first generator.

Withholding of capacity is prohibited by REMIT, but ACER can only detect this strategic behavior by carrying out a case-by-case analysis, as explained in (2.1.2). Such detection can be particularly difficult when operators, instead of physically withholding their capacity, perform economic withholding, which also results in the removal from the market of capacity that, if offered at an appropriate price, would set the price at a lower level.



γ_1	γ_2	γ_3	Q_1^*	Q_2^*	Q_3^*	λ^*
1	1	0	20	5	0	10

γ_1	γ_2	γ_3	Q_1^*	Q_2^*	Q_3^*	λ^*
1	0	1	20	0	10	15

Figure 4.7: Comparison between the optimal solutions without (left) and with (right) the withholding strategy: in order to increase the profit (green area) of the first generator and thus cover additional (e.g. fixed) costs and satisfy the RA constraint, the CP takes the second generator off-line ($\gamma_2 = 0$, orange) to force the MO to set a higher clearing price

As the offer price is a parameter of the model rather than a variable, economic withholding is not possible in our model, but the effect on the clearing price is the same as physical withholding. The constraints introduced to avoid unwanted physical withholding will be presented later in the complete model, i.e. (5.10)-(5.13).

4.1.4 RES offer at LCOE: their offer quantity can be partially accepted

Renewable cannibalization is a phenomenon observed in the energy sector, where the increasing integration of zero marginal cost renewable energy sources, such as solar and wind power, leads to a decline in their market value. This decline poses a threat to investment incentives in renewable technologies. Studies on cannibalization in high wind and solar penetration markets have been conducted for California [Prol et al., 2020] and Germany [Liebensteiner and Naumann, 2022]. These papers use a time-series econometric model using ex-post data to find yearly cannibalization results and then apply regression models to find relations between the market price and the so-called cannibalization factor and relevant factors.

The following key points outline the characteristics and implications of renewable cannibalization:

- **Zero Marginal Cost Nature**: Solar and wind power exhibit zero marginal costs once the infrastructure is established, distinguishing them from conventional power sources such as gas, coal fired ones.
- **Market Value Decline**: Increased penetration of renewable energy can lead to a decrease in electricity prices during periods of high renewable generation, impacting the market

value of renewables.

- **Impact on Investment Incentives**: The reduced market value can deter investors, as the revenue generated may not cover the costs of installing and maintaining renewable infrastructure.

- **Supply and Demand Dynamics**: The extent of cannibalization is influenced by the balance between supply (renewable generation) and demand (electricity consumption), according to the principles of supply and demand.

- **Time-Dependent Nature**: The intermittency of renewable energy, especially wind and solar, contributes to cannibalization. Solar energy is more predictable due to consistent radiation, while wind power is less predictable.

- **Geographical and Regulatory Variations**: Cannibalization research is often specific to a country or region, considering local factors like energy demand, grid infrastructure, and regulatory frameworks.

- **Technology-Specific Considerations**: Cannibalization affects solar and wind power differently. Solar is exposed to prices during the day, while wind power generation extends beyond daylight hours, influenced by night-time wind speeds

Understanding these dynamics is crucial for policymakers and investors to design effective strategies that support renewable energy growth while addressing challenges associated with cannibalization.

It is well known that at the time of writing in the day ahead market, the RES generators offer at 0 €/MWh as the market-clearing price is typically set by gas power plants. Projecting us to 2040 and beyond, we will introduce a conceptual paradigm shift in the offering strategies of RES producers which may materialize in some respect. Specifically, we will assume that RES generators offer at their, perceived - Levelized Cost of Electricity (LCOE), e.g. [Ueckerdt et al., 2013]. In its several variants, LCOE is a metric used to assess the lifetime cost of generating a unit of electricity from a particular energy generation source and is expressed in terms of the cost per MWh. LCOE is often used to compare the economic competitiveness of different energy technologies but, reversing the perspective, can be also used as a metric for offering in future systems with high penetration of RES. This is particularly relevant when payments outside the markets are excluded or only partially considered for this class of production plants.

Non-programmable renewable power sources offer on the market their entire production due to the inability to modulate their output. The production is determined by their capacity, represented by $\overline{Q}^V$ for photovoltaic generators, and the corresponding capacity factor, cf^V, influenced by solar irradiation at that specific time.

Let's consider two hypotheses for modeling the acceptance criteria of RES offers:

a. the RES offer can be partially accepted

b. the RES offer is fully accepted

We propose to operate under hypothesis (a.) so that the clearing price can coincide with the photovoltaic offer price P^V ($= LCOE^V$). Conversely, under the alternative hypothesis (b.), the LCOE will never fix the clearing price, as illustrated below.

In the case of fully accepted RES offers (b.), we introduce a slack variable OG representing RES curtailment (i.e. an overgeneration) in the balance constraint (4.15c) to ensure problem feasibility.

$$\min_{Q^K \geq 0, OG \geq 0} P^K Q^K + P^V cf^V \overline{Q}^V \tag{4.15a}$$

$$\underline{Q}^K \gamma \leq Q^K \leq \overline{Q}^K \gamma \tag{4.15b}$$

$$Q^K + cf^V \overline{Q}^V = D + OG \qquad (\lambda) \tag{4.15c}$$

The term $P^V cf^V \overline{Q}^V$ in the objective function (4.15a) is constant and can be omitted without affecting the optimal solution. The clearing price takes a value of zero ($\lambda = 0$) when there is an excess of production over demand ($OG > 0$): when satisfying the demand D, the cost of producing the entire quantity $cf^V \overline{Q}^V$ is already taken into account and one more unit of demand would be satisfied by reducing OG without changing the cost. Note that penalizing the variable OG with the photovoltaic offer price P_V in the objective function would lead P_V to be able to set the clearing price. However, this solution is limited to a single type of RES offering at a fixed price and is not applicable when dealing with multiple types of RES production (e.g. wind) offered at different prices.

Therefore, we consider the possibility of partially accepted RES offers (a.) as in (4.16).

$$\min_{Q^K \geq 0, Q^V \geq 0} P^K Q^K + P^V Q^V \tag{4.16a}$$

$$\underline{Q}^K \gamma \leq Q^K \leq \overline{Q}^K \gamma \qquad (\underline{\mu}^K, \overline{\mu}^K \geq 0) \tag{4.16b}$$

$$Q^V \leq cf^V \overline{Q}^V \qquad (\mu^V \geq 0) \tag{4.16c}$$

$$Q^K + Q^V = D \qquad (\lambda) \tag{4.16d}$$

The example in Tab. 4.2 illustrates the comparison between the models (4.15) and (4.16), corresponding to the hypotheses (b.) and (a.) respectively. The system configuration includes one

	off.price	off.quantity				
				D	10	40
K	94	[30,40]		Q^K	0	30
V	64	(a.) [0,20]	(a.)	Q^V	10	10
		or (b.) 20	(b.)	OG	10	10
			(a.)	λ	64	64
			(b.)	λ	0	0

Table 4.2: Comparison between optimal solutions under the hypotheses (a.) partially accepted and (b.) fully accepted for RES offers at different levels of demand D, given the offer prices and quantities of the thermal power plant K and the RES power plant V: under (a.), V is the marginal generator and sets the price, while under (b.) the price is zero since the cost of meeting an additional unit of demand does not vary due to the overgeneration caused by the non-partitionable quantity of V and the technical minimum of K

thermal generator with offer price 94 €/MWh and offer quantity 40 MW and one photovoltaic generator with offer price 64 €/MWh, based on its LCOE, and offer quantity 20 MW, given by the product of its capacity and capacity factor; the thermal generator presents a technical minimum level of 30 MW. The clearing price λ takes the LCOE value of 64 only under the hypothesis (a.). With $D = 40$ MWh, the photovoltaic is the marginal generator due to the technical minimum of the thermal generator.

Therefore we carry on the problem (4.16) whose dual is the problem (4.17).

$$\max_{(\underline{\mu}^K,\overline{\mu}^K,\mu^V)\geq 0,\lambda} D\lambda + \underline{Q}^K\gamma\underline{\mu}^K - \overline{Q}^K\gamma\overline{\mu}^K - cf^V\overline{Q}^V\mu^V \tag{4.17a}$$

$$-\underline{\mu}^K + \overline{\mu}^K - \lambda \geq -P^K \qquad (Q^K \geq 0) \tag{4.17b}$$

$$\mu^V - \lambda \geq -P^V \qquad (Q^V \geq 0) \tag{4.17c}$$

The CSC (4.18) are

$$\mu^V(cf^V\overline{Q}^V - Q^V) = 0 \qquad Q^V(\mu^V - \lambda + P^V) = 0 \tag{4.18}$$

and introduce the following criteria

$cf^V = 0$	$Q^{V*} = 0$	$\mu^{V*} \geq 0$	$\mu^{V*} \geq \lambda^* - P^V$
$cf^V > 0$	$Q^{V*} = 0$	$\mu^{V*} = 0$	$\lambda^* \leq P^V$
	$0 < Q^{V*} < cf^V \overline{Q}^V$	$\mu^{V*} = 0$	$\lambda^* = P^V$
	$Q^{V*} = cf^V \overline{Q}^V$	$\mu^{V*} = \lambda^* - P^V \geq 0$	$\lambda^* \geq P^V$

For $cf^V > 0$ the dual variable μ^{V*} represents the **profit per unit** of the photovoltaic generator:

$$\mu^{V*} = \max\{0, \lambda^* - P^V\} \tag{4.19}$$

For $cf^V = 0$, on the other hand, the value of μ^{V*} does not impact the objective function (4.11a) and can take on any value that is $\geq \lambda^* - P^V$. In this case, it does not represent the profit per unit.

4.1.5 Batteries to reduce energy losses

We integrate batteries into the system to efficiently manage RES production that may not be fully accepted by the MO, as described in (4.1.4). The exceeding RES production can be stored in batteries and be used in later hours for reducing thermal production and thus lowering the cost of meeting the demand realizing a so-called *time shifting*.

The CP can decide optimal investments and operation of batteries. From the operational point of view, the variables B_t^C and B_t^D are introduced in the UL. These variables, representing the charge and the discharge of the battery, can be modeled as SOS1 to establish their mutual exclusivity, i.e. to ensure that charging and discharging do not occur simultaneously. In addition, continuous variables B_t are used to represent the energy level of the battery in hour t. UL constraints model the operation of batteries. Specifically, energy balances (4.20) ensure that the energy stored at the end of hour t equals the energy stored at the end of hour $t-1$, plus the energy charged or minus the energy discharge. Additional constraints will be introduced later.

$$B_t = B_{t-1} + B_t^C - B_t^D \tag{4.20}$$

We consider two alternative formulations for the LL, that have respective different formulations for the UL:

B1 batteries not participating in the market

B2 batteries participating in the market

In the first case (B1), only the discharge B_t^D is considered as a parameter of the LL balance constraint (4.21d).

$$\min_{Q_t^K \geq 0, Q_t^V \geq 0} P^K Q_t^K + P^V Q_t^V \tag{4.21a}$$

$$\underline{Q}^K \gamma_t \leq Q_t^K \leq \overline{Q}^K \gamma_t \qquad (\underline{\mu}_t^K, \overline{\mu}_t^K \geq 0) \tag{4.21b}$$

$$Q_t^V \leq cf_t^V \overline{Q}^V \qquad (\mu_t^V \geq 0) \tag{4.21c}$$

$$Q_t^K + Q_t^V = D_t - B_t^D \qquad (\lambda_t) \tag{4.21d}$$

The charge B_t^C is limited in the UL constraint by the RES production net of the quantity accepted by the MO, e.g. for photovoltaic these constraints read as in (4.22).

$$B_t^C \leq cf_t^V \overline{Q}^V - Q_t^V \tag{4.22}$$

It's important to note that this constraint may not always be satisfied at equality due to technical limitations, such as the inability to store all excess energy. For instance, the excess energy might exceed the charging rate or the battery may already be saturated. In such cases, there could be an incentive for the CP to invest in additional battery capacity. Indeed, the model chooses the number of batteries that represents the best trade-off between the increase in investment costs for installing new batteries and the reduction in operating costs due to increased storage capacity, resulting in lower residual demand to satisfy.

From the point of view of the e.g. photovoltaic, and more generally RES, generator, the quantity Q_t^V accepted by the MO generates a revenue at the hourly price λ_t whereas the excess energy that is used to charge the battery does not produce a revenue. The described operation of the system is only possible in the presence of a supporting mechanism that compensates for the investment and service of storing and shifting the production from one hour to another.

In the second case (B2), instead, both charge B_t^C and discharge B_t^D are considered as parameters of the LL balance constraint (4.23d). Therefore, the accepted quantities incorporate the energy needed to charge the battery, which in this case can be charged by both RES and thermal generators.

$$\min_{Q_t^K \geq 0, Q_t^V \geq 0} P^K Q_t^K + P^V Q_t^V \tag{4.23a}$$

$$\underline{Q}^K \gamma_t \leq Q_t^K \leq \overline{Q}^K \gamma_t \qquad (\underline{\mu}_t^K, \overline{\mu}_t^K \geq 0) \tag{4.23b}$$

$$Q_t^V \leq cf_t^V \overline{Q}^V \qquad (\mu_t^V \geq 0) \tag{4.23c}$$

$$Q_t^K + Q_t^V \; = \; D_t + B_t^C - B_t^D \qquad (\lambda_t) \tag{4.23d}$$

RES also receive the clearing price for the part of their production used to charge the battery. Therefore in this second case, (B2), the batteries can be eligible for RA treatment.

4.1.6 The need for a fully-adequate model with side payments

As described in (1.3), energy-only markets consider revenues only from the market, relying on scarcity rents of the hours when demand exceeds supply ($ENP > 0, \lambda = VoLL$) to cover fixed and investment costs. Therefore, a CP aiming to meet RA constraints may, or is forced to, intentionally create a situation where the system cannot fulfill the entire demand. At the same time, the CP must reduce to a minimum the number of hours in which some consumers are not supplied due to system adequacy issues; in particular, an electricity system is typically assumed to be adequate if there are no more than 3 hours per year with positive ENP (*system adequacy condition*).

As explained in (4), for computational tractability, we limit our analysis to a small number of days representative of the future year. This limitation makes it difficult to restrict the number of hours when the price is set to the VoLL as per the system adequacy target. We therefore penalize the ENP in the UL objective function at the VoLL.

We observed that the CP determines optimal values of investments, statuses of thermal generators, and discharges of batteries so that in some hours the total capacity is *slightly less* than demand (small ENP values), forcing the MO to assign a price equal to the VoLL for the entire dispatched quantities.

Therefore, it is not possible to model an energy-only market with RA constraints as the latter would affect the optimal solution.

A *fully-adequate* system is considered, i.e., a system without ENP. Therefore, the clearing price can be at most equal to the maximum bid price. Defining $\hat{\lambda}_i = \{P^{RES}, P^K\}$ as the ordered set of offer prices, the inequality (4.24) holds.

$$\lambda \leq \overline{\lambda} = \max_i \hat{\lambda}_i \tag{4.24}$$

At the same time, thermal generators are assumed to bid at their marginal costs. Side payments become necessary to cover any costs beyond the marginal ones, particularly for the marginal generator, as the other can rely on infra-marginal rents. The model is referred to as "*simil-capacity*" because of the possibility for generators to receive an additional (and discriminatory) revenue, e.g. u_k for thermal power plant k, outside of the market. This side payment will be minimized by the CP and will act as a slack variable possibly needed to satisfy the RA constraints. These can be modeled with constraints that, for e.g. thermal generators, take the generic form (4.25).

$$\sum_t \lambda_t Q_{k,t} + u_k \geq \sum_t (C_k^{SU} \alpha_{k,t} + C_k^M Q_{k,t}) + \text{InvCosts} \tag{4.25}$$

The RA constraints are expressed as inequalities with slack variables, e.g. u_k, defined as positive and minimized in the objective function. In fact, in cases where market revenues, such as $\sum_t \lambda_t Q_{k,t}$ for thermal generators, exceed the associated costs, equalities would not be feasible. Modeling u_k as a free variable and removing it from the objective function would introduce distortions in market results. In fact, minimizing the price as in (5.45) would lead the CP to strategically dispatch thermal plants at their technical minimums to avoid them setting the price. The price would instead be set by less expensive generators; any losses incurred would be covered by u_k, which would not affect the value of the objective function.

4.1.7 Price indeterminacy

In the LL model, the clearing price λ_t is indeterminate when the following conditions occur simultaneously in hour t:

1. RES generators are dispatched at the maximum hourly production ($Q_t^V = cf_t^V \overline{Q}^V$).
2. ON thermal generators ($\gamma_{k,t} = 1$) are dispatched at their capacity ($Q_{k,t}^K = \overline{Q}_k^K$);
3. Batteries playing with charge and discharge to meet on a vertical line the demand, i.e. (**I3**)

The clearing price is defined within an interval of values $[\lambda^*_{min}, \lambda^*_{max}]$ as described in (**E4**) and (**I3**). Among the prices in this interval, the CP chooses one that satisfies the RA constraints, adopting an optimistic position as explained in (3.2). In Appendix A of [Caramanis and Associates Inc., 2002] the price-indeterminacy is considered an unlikely solution to the hourly auction problem and one method is suggested to deal with indeterminacy. However, this approach does not solve the issue in this situation. On the other hand, the model can strategically identify investment decisions that are not realistic. In fact, the CP can identify solutions that combine the operational decision of commitment of the thermal generator and the investment decision in RES capacity in such a way that the total quantity offered coincides with the quantity demanded, and the price becomes indeterminate. As we shall see soon, from the CP standpoint choosing the upper bound of the price does not change its objective function while it helps in satisfying the RA constraints.

Since in reality such behavior on RES, w.r.t the continuous decisions, cannot be applied by investors, we must eliminate these types of solutions by modeling discrete investment decisions (lumpy investments). The constraints introduced to avoid unwanted price indeterminacy created by occurrences (2)-(3) will be presented later in the complete model, i.e. (5.25)-(5.30).

4.2 Models description

Let us consider the Generation Expansion Planning problem determining investments in new thermal plants, new RES plants (photovoltaic and wind power), and new batteries so that the resulting mix is able to meet future electricity demand as the first goal. A simplified transportation model is used to represent an electricity system consisting of a set of **zones** interconnected by transmission **lines** of limited capacity. We assume that during the considered planning horizon, there are no changes in the transmission topology, i.e., the existing transmission lines remain the same or are updated accordingly with an off-line process.

To balance modeling accuracy and computational complexity, we develop a **static** GEP model, considering a future year (the year 2040 in our applications to the Italian system) that is represented by a small number of **representative days** chosen in such a way as to take into account the most significant and relevant operating conditions of the system, as detailed in (4). The dynamic approach could provide more accurate solutions but it would increase the complexity of the problem, as detailed in (2).

The **initial state of the system** is represented by a set of existing thermal and renewable plants and a set of existing batteries. We do not consider specific thermal plant decommissioning; we will assess ex-post which plants are not chosen to meet the load. Additionally the model could intercept policy targets by simply setting to certain maximum values the usage of specific technologies, such as (initially existing) coal.

The new production mix must be determined taking into account the **targets for decarbonization** of energy systems: we have in particular considered that in each zone z:

- the final installed capacity (existing + new) of photovoltaic power generation must be at least equal to the minimum required capacity
- the final installed capacity (existing + new) of wind power generation, considering both

in-shore and offshore, must be at least equal to the minimum required capacity.

In a system with high shares of renewable capacity, thermal plants only produce when the available primary sources (solar radiation, wind speed) are insufficient to meet the demand. Thermal plants are therefore used for a reduced number of hours, w.r.t. the classic base load operation, and likely incur higher operating costs due to the increased cycling operations needed to cope with the non-programmability of renewable plants. From the modeling standpoint, the thermal power plants dynamics in tactical, short-term, problems are defined with typical constraints such as Technical Minimum power, Minimum Up and Down times, and max Up and Down Ramps rates. As for the costs, the models typically include also Start Up Costs that are incurred whenever such plants are switched on and that can depend on how long the plant has been off-line, e.g. [Van Ackooij et al., 2018], [Gentile et al., 2017]. On the contrary in strategic models such as the ones of this work, it is custom to neglect some of these since their inclusion becomes unjustified when considering the uncertainty and if compared with the additional computational burden. For instance, in [Poncelet et al., 2020] it is argued that if storage technologies are considered, integrating technical constraints for thermal plants has only a minor impact on both overall cost projections and most investments. Lastly, it is important to point out that especially for the new gas-fired plants, manufacturers are reducing the physiological values of the Min Up and Down times and enlarging the ramp rate capabilities. For economic reasons in the forthcoming models, we decided to keep the Start Up Costs since they are more and more entering into the economic evaluation of the profitability of such plants. Apart from the technical constraints in the proposed models we impose economic constraints related to the **revenue adequacy of thermal plants** over the period considered, also for the reasons explained in (1.3). Later we will also explain how one component of these constraints can be interpreted with the lens of the CRM.

In order to correctly estimate the revenues of thermal plants, we want the zonal hourly price to be defined as *the cost of satisfying an additional unit of demand*. The zonal hourly price that is endogenously determined by our model meets the definition and is not affected by the RA constraint, thus solving the problems inherent in the approach [Guo et al., 2022] (see 2.2.2).

In fact, we develop a **bilevel model** in which

- the lower level (LL) is the problem solved for each hour by the Market Operator (MO): given (a) the hourly-zonal demand (inelastic), (b) the supply bids submitted by individual generators, and (c) the operation of the batteries, the MO determines the dispatch of the generators, i.e. the accepted quantities of each bid; in such a model, the hourly-zonal clearing price is the optimal value of the dual variable associated with the constraint imposed to satisfy the hourly-zonal demand.

- the upper level (UL) is the problem of a central entity (Central Planner - CP), which
  - determines the evolution of the system, i.e. the investments in thermal plants, RES plants, and batteries
  - determines the operation of thermal power plants and batteries to meet the hourly-zonal demand.

  The CP combines in itself the role performed by various actors in the system (investors, generation plant operators, battery operators, and the regulator).

In the system, we therefore consider one main decision maker, the CP, and a price-setting entity, the MO, that responds to the actions of the CP, giving birth to a bilevel optimization model.

As described in (4.1.6) it is assumed that:

- the system is *fully adequate*, i.e., there is no Energy Not Provided and the clearing price can be at most equal to the maximum bid price (no VoLL)
- thermal generators bid at marginal cost.

Therefore **side-payment** may be required to cover fixed and investment costs.

Focusing on the final goal of the GEP models it is clear that we seek an "optimal" resource mix. However, the mathematical formulation of "optimal" in the objective functions is fuzzy and can be differently declined. For instance, one may focus on the total operational costs as in traditional models, or it may focus on system costs which may differ from operational costs. Additionally, it can include other components such as side payments. Moreover, the investment costs can appear in the OF or somewhere else. As a general consideration, we can therefore state that GEP models are not universally defined and modeling choices have to be made.

We proposed two alternative models, which can be conceptually sketched as in Fig. 4.8.

1. GEP-RA (1)
2. GEP-RA (2)

The colors indicate the type of variables: red for continuous variables, green for integer (binary) variables, and blue for dual variables associated with the constraints of the LL problem. Both problems have a hierarchical structure as the UL problem includes the optimal solution of the LL problem but not vice versa. Instead, in the LL problem, the UL variables are fixed parameters and not decision variables. The constraints on integer decisions, e.g., the on/off

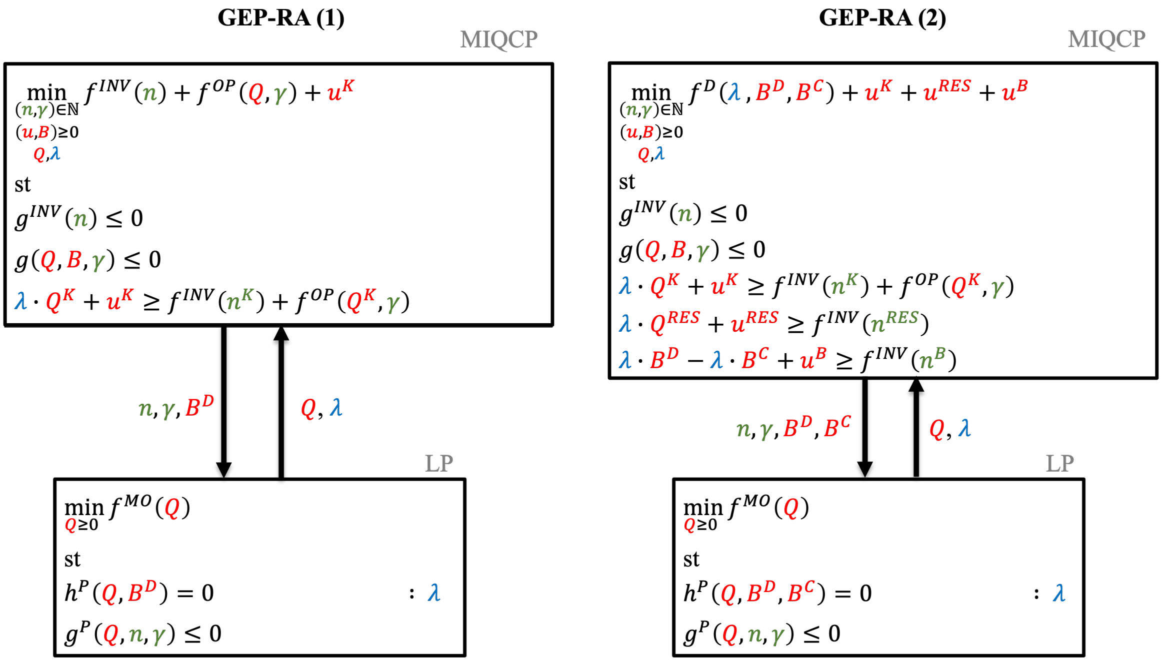


Figure 4.8: Proposed bilevel models: the lower levels determine the accepted quantities Q and the market clearing prices λ given the investments n and the operational decisions γ, B^D, B^C taken by the upper levels; GEP-RA (1) minimizes both the investment and operational costs taking into account the revenue adequacy constraints only for thermal power plants, whose side payments u are also minimized, while GEP-RA (2) minimizes the system costs additionally taking into account the revenue adequacy of renewables and batteries by introducing battery charging in the lower level

statuses γ of conventional generators, are enforced in the UL problem and the corresponding integer decisions parameterize the LL problems so that the LL is linear. Similarly, the dispatch decisions, e.g., the accepted quantities Q^K of conventional generators, resulting from the LL problems affect the decisions made in the UL problem. The LL problem yields clearing prices λ, which, in turn, are used in the UL problem to compute the revenues collected, e.g., by conventional generators on the market. RA constraints make the UL a mixed-integer quadratically constrained program.

In GEP-RA (1), following the conventional approach of centralized GEP models (2.2.1), the investment $f^{INV}(n)$ and operational costs $f^{OP}(Q,\gamma)$ are minimized in the UL. RA constraints are considered for thermal generators in a framework akin to the Italian Capacity Market, with side payments u^K also minimized in the OF. In GEP-RA (2), on the other hand, the system costs (i.e. the total costs to consumers) are minimized, moving $f^{INV}(n)$, $f^{OP}(Q,\gamma)$ from the OF to the RA constraints, since these costs are borne by the generators' owners and not by the system. In the context of the FERX consultation, RA constraints are introduced also for RES and batteries to cover their investment costs, with specific side payments minimized in the OF. To account for the revenues of batteries, GEP-RA (2) considers a LL with both charge and discharge as described in (B2).

Chapter 5

The proposed models

In this chapter, we provide a detailed formulation of the models GEP-RA (1) and GEP-RA (2) introduced in Section (4.2). Furthermore, enhancements to certain constraints are proposed in Section (5.3), and a version of the model considering only continuous and binary variables is provided in Section (5.4).

5.1 GEP-RA (1)

The GEP-RA (1) model is (5.1)-(5.39). Both the Market Operator (MO) model and the Central Planner (CP) model are included.

$$\min_{Var} TC \tag{5.1}$$

Variables

$Var \equiv$

$\left\{n_z^V, n_z^{WI}, n_z^{WO}, n_z^B, n_k^K\right\}$ (Investments) $\cup$

$\left\{u_k^K\right\}$ (Side payments) $\cup$

$\left\{\alpha_{k,t}, \beta_{k,t}, \gamma_{k,t}, \gamma_{k,t}^{SL}, x_{k,t}^{TF}\right\}$ (Unit commitment with avoidance of capacity withholding) $\cup$

$\left\{B_{z,t}, B_{z,t}^C, B_{z,t}^D, \phi_{z,t}\right\}$ (Batteries) $\cup$

$\{RNP_{z,t}\}$ (Reserve not provided) $\cup$

$\{y_{z,t,i}\}$ (Price indeterminacy) $\cup$

$$\left\{Q_{k,t}^K, Q_{z,t}^V, Q_{z,t}^{WI}, Q_{z,t}^{WO}, \lambda_{z,t}\right\} \text{ (LL - optimistic perspective)} \tag{5.2}$$

Objective function

$$
\begin{aligned}
&TC = \\
&= \frac{|T|}{8760} \sum_{z \in Z} \Big(I_z^V \ \overline{Q}^V \ n_z^V + I_z^{WI} \ \overline{Q}^{WI} \ n_z^{WI} + I_z^{WO} \ \overline{Q}^{WO} \ n_z^{WO} + I_z^B \ \overline{Q}^B \ n_z^B \Big) + \\
&+ \frac{|T|}{8760} \sum_{k \in K} I_k^K \ \overline{Q}_k^K \ n_k^K + \sum_{k \in K} u_k^K + \sum_{k \in K} \sum_{t \in T} (C_k^{SU} \ \alpha_{k,t} + C_k^M \ Q_{k,t}^K) + \\
&+ \sum_{z \in Z} \sum_{t \in T} c^{RNP} \ RNP_{z,t}
\end{aligned} \tag{5.3}
$$

Constraints on zonal capacities of RES power plants and batteries in the target year: $z \in Z$

$$\underline{V}_z \leq \overline{Q}^V \ (n_{z,0}^V + n_z^V) \leq \overline{V}_z \tag{5.4}$$

$$\underline{W}_z \leq \overline{Q}^{WI} \ (n_{z,0}^{WI} + n_z^{WI}) + \overline{Q}^{WO} \ (n_{z,0}^{WO} + n_z^{WO}) \leq \overline{W}_z \tag{5.5}$$

$$\underline{B}_z \leq \overline{Q}^B \ (n_{z,0}^B + n_z^B) \leq \overline{B}_z \tag{5.6}$$

Constraints on thermal power capacity in the target year: $k \in K_z, \ z \in Z$

$$n_k^K \leq \overline{n}_k^K \tag{5.7}$$

UC constraints, with avoidance of physical withholding

$$\gamma_{k,t} - \gamma_{k,t}^I = \alpha_{k,t} - \beta_{k,t} \qquad k \in K, t \in T^I \tag{5.8}$$

$$\gamma_{k,t} - \gamma_{k,t-1} = \alpha_{k,t} - \beta_{k,t} \qquad k \in K, t \in T \setminus T^I \tag{5.9}$$

$$\gamma_{k,t} + \gamma_{k,t}^{SL} = n_{k,0}^K + n_k^K \qquad k \in K, t \in T \tag{5.10}$$

$$\gamma_{k,t}^{SL} \geq 1 - x_{k,t}^{TF} \qquad k \in K, t \in T \tag{5.11}$$

$$\gamma_{k,t}^{SL} \leq (n_{k,0}^K + \overline{n}_k^K)(1 - x_{k,t}^{TF}) \qquad k \in K, t \in T \tag{5.12}$$

$$x_{k,t}^{TF} \geq x_{kk,t}^{TF} \qquad q \in [1, |Q|-1], k \in K_q, kk \in K_{q+1}, t \in T \tag{5.13}$$

Constraints that define accepted quantities and zonal prices for each hour: $t \in T$

Optimality conditions of the MO problem (5.42) (5.14)

Constraints on operation of batteries: $z \in Z$

$$B^C_{z,t} \leq rt^C\ \overline{B}_z\ \phi_{z,t} \qquad t \in T \quad (5.15)$$

$$B^D_{z,t} \leq D_{z,t}\ (1 - \phi_{z,t}) \qquad t \in T \quad (5.16)$$

$$\begin{aligned} B^C_{z,t} \leq\ & cf^V_{z,t}\ \overline{Q}^V\ (n^V_{z,0} + n^V_z) + cf^{WI}_{z,t}\ \overline{Q}^{WI}\ (n^{WI}_{z,0} + n^{WI}_z) \\ & + cf^{WO}_{z,t}\ \overline{Q}^{WO}\ (n^{WO}_{z,0} + n^{WO}_z) - Q^V_{z,t} - Q^{WI}_{z,t} - Q^{WO}_{z,t} \end{aligned} \qquad t \in T \quad (5.17)$$

$$B^C_{z,t} \leq rt^C\ \overline{Q}^B\ (n^B_{z,0} + n^B_z) \qquad t \in T \quad (5.18)$$

$$B^D_{z,t} \leq rt^D\ \overline{Q}^B\ (n^B_{z,0} + n^B_z) \qquad t \in T \quad (5.19)$$

$$B_{z,t} \leq \overline{Q}^B\ (n^B_{z,0} + n^B_z) \qquad t \in T \quad (5.20)$$

$$B_{z,t} = (1 - \kappa)\ B^I_{z,t} + \kappa^C\ B^C_{z,t} - \kappa^D\ B^D_{z,t} \qquad t \in T^I \quad (5.21)$$

$$B_{z,t} = (1 - \kappa)\ B_{z,t-1} + \kappa^C\ B^C_{z,t} - \kappa^D\ B^D_{z,t} \qquad t \in T \setminus T^I \quad (5.22)$$

$$B_{z,t} = B^I_{z,t} \qquad t \in T^L \quad (5.23)$$

Reserve constraints: $z \in Z, t \in T$

$$\begin{aligned} & \sum_{k \in K_z} (\overline{Q}^K_k\ \gamma_{k,t} - Q^K_{k,t}) + \\ & + df\left[rt^D\ \overline{Q}^B\ (n^B_{z,0} + n^B_z)\ (1 - \phi_{z,t}) - B^D_{z,t}\right] + RNP_{z,t} \geq R_{z,t} \end{aligned} \quad (5.24)$$

Constraints that avoid price indeterminacy: $z \in Z,\ t \in T$

$$y_{z,t,1} \leq \sum_{z' \in Z} Q^V_{z',t} \quad (5.25)$$

$$y_{z,t,2} \leq \sum_{z' \in Z} Q^{WI}_{z',t} \quad (5.26)$$

$$y_{z,t,3} \leq \sum_{z' \in Z} Q^{WO}_{z',t} \quad (5.27)$$

$$y_{z,t,3+q} \leq \sum_{k \in K_q} \gamma_{k,t} \qquad q \in Q \quad (5.28)$$

$$\sum_{i \in I} y_{z,t,i} = 1 \quad (5.29)$$

$$\lambda_{z,t} = \sum_{i \in I} \hat{\lambda}_i\ y_{z,t,i} \quad (5.30)$$

Revenue adequacy constraints: $k \in K_z, \ z \in Z$

$$\sum_{t \in T} \lambda_{z,t} \ Q^K_{k,t} + u^K_k \geq M \left[\frac{|T|}{8760} \ I^K_k \ \overline{Q}^K_k \ n^K_k + \sum_{t \in T} (C^M_k \ Q^K_{k,t} + C^{SU}_k \ \alpha_{k,t}) \right] \tag{5.31}$$

Variables types

$$n^V_z, n^{WI}_z, n^{WO}_z, n^B_z \in \mathbb{N} \qquad z \in Z \tag{5.32}$$

$$n^K_k \in \mathbb{N} \qquad k \in K \tag{5.33}$$

$$u^K_k \geq 0 \qquad k \in K \tag{5.34}$$

$$\alpha_{k,t}, \beta_{k,t}, \gamma_{k,t}, \gamma^{SL}_{k,t} \in \mathbb{N} \qquad k \in K, t \in T \tag{5.35}$$

$$x^{TF}_{k,t} \in \{0,1\} \qquad k \in K, t \in T \tag{5.36}$$

$$B_{z,t}, B^C_{z,t}, B^D_{z,t}, RNP_{z,t} \geq 0 \qquad z \in Z, t \in T \tag{5.37}$$

$$\phi_{z,t} \in \{0,1\} \qquad z \in Z, t \in T \tag{5.38}$$

$$y_{z,t,i} \in \{0,1\} \qquad z \in Z, t \in T, i \in I \tag{5.39}$$

5.1.1 Capacity of RES power plants, thermal power plants and batteries in the target year

We assume standard sizes for RES power plants and batteries, expressed by the following parameters

$\overline{Q}^V$	$[MW]$	Standard capacity of solar power plants
$\overline{Q}^{WI}$	$[MW]$	Standard capacity of onshore wind power plants
$\overline{Q}^{WO}$	$[MW]$	Standard capacity of offshore wind power plants
$\overline{Q}^B$	$[MW]$	Standard capacity of batteries

Thermal power plants can have different capacities. The types of thermal power plants located in zone $z \in Z$ are the elements of set K_z and the different capacities are represented by the following parameters

$\overline{Q}^K_k$	$[MW]$	Capacity of thermal power plants of type $k \in K_z$ for $z \in Z$

The configuration of the system at the beginning of the planning period is represented by the following parameters

$n^V_{z,0}$	[−]	Number of solar power plants (standard capacity) existing in zone z at the beginning of the planning period
$n^{WI}_{z,0}$	[−]	Number of onshore wind power plants (standard capacity) existing in zone z at the beginning of the planning period
$n^{WO}_{z,0}$	[−]	Number of offshore wind power plants (standard capacity) existing in zone z at the beginning of the planning period
$n^B_{z,0}$	[−]	Number of batteries (standard capacity) existing in zone z at the beginning of the planning period
$n^K_{k,0}$	[−]	Number of thermal power plants of type k existing in zone z at the beginning of the planning period

In each zone z, the CP representing the investors determines the number of new RES power plants, new thermal power plants and new batteries represented by the following integer-valued variables

n^V_z	[−]	Number of new solar power plants in zone z in the target year
n^{WI}_z	[−]	Number of new onshore wind power plants in zone z in the target year
n^{WO}_z	[−]	Number of new offshore wind power plants in zone z in the target year
n^B_z	[−]	Number of new batteries in zone z in the target year
n^K_k	[−]	Number of new thermal power plants of type k in the target year

Constraints (5.4), (5.5), and (5.6) impose that in the target year, the total capacity of solar power plants, wind power plants, and batteries, respectively, in each zone z is between the following lower and upper bounds, which represent policy requirements and budget/permitting constraints, respectively.

$\underline{V}_z$	$[MW]$	Minimum solar power capacity in zone z in the target year
$\overline{V}_z$	$[MW]$	Maximum solar power capacity in zone z in the target year
$\underline{W}_z$	$[MW]$	Minimum wind power capacity in zone z in the target year
$\overline{W}_z$	$[MW]$	Maximum wind power capacity in zone z in the target year
$\underline{B}_z$	$[MW]$	Minimum storage capacity in zone z in the target year
$\overline{B}_z$	$[MW]$	Maximum storage capacity in zone z in the target year

Constraints (5.7) impose that in the target year, the number of thermal power plants of each type k in each zone z is bounded above by $\overline{n}_k^K$.

5.1.2 Unit commitment with avoidance of physical withholding

The so-called *Clustered Unit Commitment formulation* is used to take into account Unit Commitment constraints in the optimization model while keeping the problem computationally tractable, as described in (5). Although simplified, the UC formulation allows for the inclusion of thermal plant startup costs over the considered planning period.

Constraints (5.8) and (5.9) ensure consistency of the values taken on in adjacent hours by the following integer variables representing startups, shutdowns, and statuses for each power plant type k.

$\alpha_{k,t}$	[−]	number of plants of type k started up at hour t
$\beta_{k,t}$	[−]	number of plants of type k shut down at hour t
$\gamma_{k,t}$	[−]	number of plants of type k on-line at hour t

Constraints (5.8) refer to the first hour (i.e., $t \in T^I$) of each representative day, for which the following parameter is given, while constraints (5.9) refer to all hours except the first (i.e., $t \in T \setminus T^I$) of each representative day.

$\gamma_{k,t}^I$	[−]	number of on-line thermal power plants of type k at the beginning of hour $t \in T^I$

Constraints (5.10) state that the number of on-line plants of type k at hour t is bounded above by the number of plants of type k installed in the target year, being $\gamma_{k,t}^{SL}$ an integer-valued, non-negative variable. In the usual UC formulation, this upper bound constraint is expressed as (5.40).

$$\gamma_{k,t} \leq n_{k,0}^K + n_k^K \qquad k \in K,\ t \in T \quad (5.40)$$

However, we have decided to integrate the UC constraints with innovative constraints, avoiding the manipulative behavior of **physical withholding of capacity** that the CP representative of the producers can adopt with the classical formulation. In fact, as described in Section (4.1.3), some thermal generators do not offer, with the aim of forcing the clearing price to be set by generators with higher marginal costs/offer prices; in this way, the producers can obtain higher revenues from other generators. The regulation prohibits this strategic behavior that is however difficult or demanding to detect.

In any case, the user of the proposed model can decide either to leave the CP free to adopt the manipulative behavior, observing ex-post from the calculated optimal solution whether such behavior is necessary to guarantee the Revenue Adequacy constraints (thanks to an increase in market revenues, see Section (5.1.7)), or impose that the CP, as the regulator, prevents such manipulative behavior for thermal generators of cluster k in hour t. From a mathematical standpoint avoiding this occurrence is quite tricky. We propose to proceed as follows: we started from the possible situation where some plant(s) in a cluster have not been offered to completion, mathematically this means that the strict inequality $\gamma_{k,t} < n^K_{k,0} + n^K_k$ holds for some k and t.

Then we introduce the integer slack variable $\gamma^{SL}_{k,t}$ into the general constraint (5.40), which now becomes (5.10).

$$\gamma_{k,t} + \gamma^{SL}_{k,t} = n^K_{k,0} + n^K_k \qquad k \in K, t \in T \quad (5.10)$$

Subsequently, we introduce a binary variable $x^{TF}_{k,t}$, linked to the integer $\gamma^{SL}_{k,t}$ according to the logical condition (5.41)

$$x^{TF}_{k,t} = \begin{cases} 1 & \text{if } \gamma^{SL}_{k,t} = 0 \\ 0 & \text{if } \gamma^{SL}_{k,t} \geq 1 \end{cases} \qquad k \in K, t \in T \quad (5.41)$$

$\gamma^{SL}_{k,t} = 0$ indicates that the generators of cluster k do not withhold capacity in hour t because all available generators (existing + built) are on-line or, in other words, present their bids ($\gamma_{k,t} = n^K_{k,0} + n^K_k$). This must trigger $x^{TF}_{k,t} = 1$. Conversely, $\gamma^{SL}_{k,t} \geq 1$ indicates that some available generators of cluster k are off-line in hour t for two possible reasons: either because such generators are not needed to fulfill the demand requirement or because they withhold capacity ($\gamma_{k,t} < n^K_{k,0} + n^K_k$). In this case, the logical constraints impose $x^{TF}_{k,t} = 0$.

We achieve our final goal by introducing precedence constraints (5.13), which for each pair of generators of clusters k and kk, differentiated by bid price ($k \in K_q$ has bid price $\hat{\lambda}_q$ lower than the bid price $\hat{\lambda}_{q+1}$ of $kk \in K_{q+1}$), require that the bids of the k generators be submitted first, i.e., $x^{TF}_{kk,t}$ cannot take the value 1 if $x^{TF}_{k,t} = 0$.

$$x^{TF}_{k,t} \geq x^{TF}_{kk,t} \qquad q \in [1, |Q| - 1], k \in K_q,\ kk \in K_{q+1}, t \in T \quad (5.13)$$

The logical condition (5.41) can be modelled with the linear constraints (5.11), (5.12), where $(n^K_{k,0} + \overline{n}^K_k)$ represents the maximum value that the variable $\gamma^{SL}_{k,t}$ can take, i.e. the maximum number of plants of cluster k.

$$\gamma^{SL}_{k,t} \geq (1 - x^{TF}_{k,t}) \qquad k \in K, t \in T \quad (5.11)$$

$$\gamma^{SL}_{k,t} \leq (n^K_{k,0} + \overline{n}^K_k)(1 - x^{TF}_{k,t}) \qquad k \in K, t \in T \quad (5.12)$$

5.1.3 Determination of accepted quantities and zonal prices

We assume that at each hour t and in each zone z the RES power plants offer their hourly production and the *on-line* thermal power plants of type k offer their capacity. The production of RES power plants depends on the respective capacity factor

$cf^V_{z,t}$	$[MWh/MW]$	Solar capacity factor of zone z in hour t
$cf^{WI}_{z,t}$	$[MWh/MW]$	Inshore wind capacity factor of zone z in hour t
$cf^{WO}_{z,t}$	$[MWh/MW]$	Offshore wind capacity factor of zone z in hour t

The offer prices are represented by the following parameters

P^V	$[€/MWh]$	Offer price of solar power production (either 0 or $LCOE^V$)
P^{WI}	$[€/MWh]$	Offer price of inshore wind power production (either 0 or $LCOE^{WI}$)
P^{WO}	$[€/MWh]$	Offer price of offshore wind power production (either 0 or $LCOE^{WO}$)
P^K_k	$[€/MWh]$	Offer price of production of thermal power plants of type k (C^M_k)

The offer price for renewable energy can either be 0, as is currently the case, or the Levelized Cost of Electricity (LCOE), according to the paradigm shift described in (4.1.4). Indeed, RES can currently rely on the clearing price set by gas-fired power plants, except for a limited number of hours per year (e.g. 100) when the clearing price is set at 0. By 2040, as RES penetration increases, the zero offer price strategy is expected to be unsustainable for RES generators, as it would set the clearing price to 0 in a much larger number of hours per year, preventing RES generators from recovering their investment costs. The offer price for thermal energy is assumed to be equal to the marginal cost C^M_k.

The market clearing problem solved by the MO for each hour $t \in T$ is as follows:

$$\min_{Q, F_{l,t}} LLP = \sum_{z \in Z} \left(P^V \, Q^V_{z,t} + P^{WI} \, Q^{WI}_{z,t} + P^{WO} \, Q^{WO}_{z,t} \right) + \sum_{k \in K} P^K_k \, Q^K_{k,t} \qquad (5.42a)$$

$$Q^V_{z,t} + Q^{WI}_{z,t} + Q^{WO}_{z,t} + \sum_{k \in K_z} Q^K_{k,t} + B^D_{z,t}$$

$$+ \sum_{l \in BS_z} F_{l,t} \geq D_{z,t} + \sum_{l \in FS_z} F_{l,t} \qquad z \in Z \qquad (\lambda_{z,t} \geq 0) \quad (5.42b)$$

$$0 \leq Q^V_{z,t} \leq cf^V_{z,t}\ \overline{Q}^V\ (n^V_{z,0} + n^V_z) \qquad z \in Z \qquad (\mu^V_{z,t} \geq 0) \quad (5.42c)$$

$$0 \leq Q^{WI}_{z,t} \leq cf^{WI}_{z,t}\ \overline{Q}^{WI}\ (n^{WI}_{z,0} + n^{WI}_z) \qquad z \in Z \qquad (\mu^{WI}_{z,t} \geq 0) \quad (5.42d)$$

$$0 \leq Q^{WO}_{z,t} \leq cf^{WO}_{z,t}\ \overline{Q}^{WO}\ (n^{WO}_{z,0} + n^{WO}_z) \qquad z \in Z \qquad (\mu^{WO}_{z,t} \geq 0) \quad (5.42e)$$

$$\underline{Q}^K_k\ \gamma_{k,t} \leq Q^K_{k,t} \leq \overline{Q}^K_k\ \gamma_{k,t} \qquad k \in K \qquad (\underline{\mu}^K_{k,t} \geq 0, \overline{\mu}^K_{k,t} \geq 0) \quad (5.42f)$$

$$\underline{F}_l \leq F_{l,t} \leq \overline{F}_l \qquad l \in L \qquad (\underline{\upsilon}_{l,t} \geq 0, \overline{\upsilon}_{l,t} \geq 0) \quad (5.42g)$$

By solving problem (5.42), the MO determines the optimal values of the following variables for hour $t \in T$:

$Q^V_{z,t}$	$[MWh]$	Accepted quantity of solar production in zone z
$Q^{WI}_{z,t}$	$[MWh]$	Accepted quantity of onshore wind production in zone z
$Q^{WO}_{z,t}$	$[MWh]$	Accepted quantity of offshore wind production in zone z
$Q^K_{k,t}$	$[MWh]$	Accepted quantity of production of thermal power plants of type k
$F_{l,t}$	$[MWh]$	Flow on transmission line l

Constraints (5.42b) state that in each zone z the sum of (i) the accepted quantities of sell offers submitted by the power plants located in the zone, (ii) the energy withdrawn from batteries located in the zone (decided by the CP), (iii) the energy imported from other zones, is at least equal to the zone's demand plus the energy exported to other zones. The optimal value of the dual variable $\lambda_{z,t}$ is the zonal clearing price. By formulating the constraint as an inequality, rather than an equality, the zonal prices take on non-negative values. Constraints (5.42c)-(5.42e) state that in each zone z the accepted quantities of RES sell offers must be non-negative and not greater than the quantity produced, which depends on the number of new generators (decided by the CP). The possibility of **partial acceptance** of RES producers' offers is considered so that the clearing price can coincide with the offer price of a RES producer; if this were not allowed, the price would be zero in case of overproduction (overgeneration) and never equal to the offer price of a RES producer, as detailed in Section (4.1.4). Constraints (5.42f) state that in each zone z the accepted quantities of thermal power plants of type k must be either within the interval defined by the technical minimum $\underline{Q}^K_k$ and the capacity $\overline{Q}^K_k$, if there are plants of type k on-line ($\gamma_{k,t} \geq 1$), or zero, if all plants of type k are off-line ($\gamma_{k,t} = 0$).

Note that $\gamma_{k,t}$, i.e., whether plants of type k present an offer or not, is determined by the CP. Constraint (5.42g) states that the flow $F_{l,t}$ must be between the minimum flow $\underline{F}_l$ and the maximum $\overline{F}_l$ flow. Among all solutions satisfying constraints (5.42b) − (5.42g), a solution must be found that minimizes the total operating cost (5.42a), i.e. the sum of the products of the accepted quantities times the corresponding offer prices.

The CP's decisions $\gamma_{k,t}$, $B^D_{z,t}$, n^V_z, n^{WI}_z and n^{WO}_z, for $k \in K_z$, $z \in Z$ and $t \in T$, are treated as parameters when deriving the *Karush-Kuhn-Tucker optimality conditions* (KKTs) of the MO problem. By including the KKTs of the MO problem in the GEP-RA (1) model, the clearing price is defined endogenously (see Section 6.1). However, the question of *price indeterminacy* is left open. Price indeterminacy occurs when demand, net of battery discharge, is equal to the sum of the capacities of the generators whose offers are accepted; in this case, the clearing price is defined in an interval of values as detailed in Section (4.1.7).

5.1.4 Operation of batteries

Constraints (5.17) to (5.23) define the following variables for the batteries located in zone z

$B_{z,t}$	$[MWh]$	energy stored at the end of hour t
$B^C_{z,t}$	$[MWh]$	charge at hour t
$B^D_{z,t}$	$[MWh]$	discharge at hour t

considering technical limits resulting from the following parameters

rt^C	$[-]$	rate on battery charge (0.25)
rt^D	$[-]$	rate on battery discharge (0.25)

and losses by means of the following parameters

κ	$[-]$	energy loss coefficient of batteries (0)
κ^C	$[-]$	loss coefficient for battery charge (1)
κ^D	$[-]$	loss coefficient for battery discharge (1)

In particular, battery charge and discharge cannot occur at the same hour t in any zone z: this is ensured by constraints (5.15) and (5.16), in which the binary variable $\phi_{z,t} \in \{0, 1\}$ is 1 if batteries are charging and 0 if batteries are discharging. In (5.16), the demand value $D_{z,t}$ is used as the upper bound, as discharging is done to meet demand. In (5.15), the value $rt^C\ \overline{B}_z$ is used as the upper bound for charging (see Tab. 6.3).

At each hour t the batteries located in each zone z may be charged using the zone's RES production that exceeds the amount accepted by the market operator. Therefore, constraints (5.17) state that the charge is bounded above by the RES production net of the quantity accepted by the MO. Constraint (5.17) may not be met at equality when it is not possible to store all the excess electricity due to the technical constraints of the battery. For example, the battery may be saturated (resulting in an incentive to invest in more battery capacity to be able to charge) or the excess energy may exceed the charge rate. The batteries are subject to the technical limits resulting from the rates rt^C and rt^D: constraints (5.18) state that in any hour t the battery charge in zone z is bounded above by the charge rate rt^C multiplied by the capacity in the target year; similarly, constraints (5.19) state that in any hour t the battery discharge in zone z is bounded above by the charge rate rt^D multiplied by the capacity in the target year. Therefore, we assume ramp values proportional to the capacity of the battery; for example, with values for rates of 0.25, each battery can complete a cycle in 4 hours. Constraints (5.20) state that the electricity stored in zone z at the end of hour t cannot exceed the storage capacity installed in the zone.

The storage balance constraints (5.21) refer to the first hour (i.e., $t \in T^I$) of each representative day, where $B^I_{z,t}$ denotes the electricity stored at the beginning of each representative day in the batteries located in zone z. The storage balance constraints (5.22) refer to all hours except the first (i.e., $t \in T \setminus T^I$) of each representative day. In both constraints, the consistency of the values taken on in adjacent hours by the energy levels is enforced.

Constraints (5.23) state that in each zone z the electricity stored at the end of the last hour of the day (i.e., $t \in T^L$) must be equal to the electricity stored at the beginning of the day, therefore enforcing a daily cycle. In this way, at the beginning of the next day, the operational situation is reconstituted as it was on the day under consideration. Constraints (5.23) prevent the model from being short-sighted (end-of-horizon effect[1]). We note that if $B_{z,t} > B^I_{z,t}$ for $t \in T^L$ (i.e., at the end of the day), the initial battery level will be higher on the following day, reducing the possibility of storing excess energy. Also, a discharge at the end of the day, resulting in $B_{z,t} < B^I_{z,t}$ for $t \in T^L$, would not be justified by a price opportunity, as we do not consider battery revenues in this model.

The above formulations for charging and discharging operations with binary variables are exact. A linear relaxation is also proposed in [Pozo, 2022], and its convex hull in [Pozo, 2023]. The author argues that only a certain percentage of overlapping charge and discharge operations

[1]The initial period is influenced by past decisions, such as the discharge or charge of batteries from previous periods. Given that the tests cannot be performed over an infinite horizon, each instance suffers from boundary effects at the end of the horizon, as no further discharge or charge releases are forecast beyond period T^L.

are observed while simplifying the model.

5.1.5 Reserve

Another relevant detail that we have included in the proposed model is the possibility of considering solutions that are not only adequate from the point of view of the electrical load, but also from the point of view of the reserve: we secure the reserve for the ancillary market (MSD), so that generators cannot be used at full capacity and can be re-dispatched to meet the system requirements. In fact, e.g. if downstream of the day-ahead and intra-day market (MGP+MI) one generator is switched on and dispatched at full capacity and one is switched off, the TSO (Terna) re-dispatches them both, calling the first one down and the second one up, in order to have, with the same total production, two generators already switched on at medium load that can be used in real-time to e.g. balance the system, resolving possible grid congestion or deal with sufficient voltage profiles.

In [Guo et al., 2022], revenues from both the sale of energy and the provision of reserve services are considered (Energy and Reserve market, typical of the US configuration). Unlike [Guo et al., 2022], we do not consider reserve constraints in the lower-level problem. In fact, in the EU, generators do not earn from reserve services in energy markets, but in separate ancillary services markets. These markets are also controlled by different entities with different criteria. For instance, in Italy, the NEMO clears the energy market with a pay-as-cleared mechanism, while the TSO clears the ancillary services market with a pay-as-bid mechanism. As we do not model the MSD, but only the MGP+MI markets, we consider the reserve constraints in the UL as a safeguard for planning purposes only; otherwise the model might decide to invest in the number of generators that, when on-line at full capacity, exactly cover the highest demand, without leaving the possibility of re-dispatching the generators to meet the system requirements, such as the tertiary reserve.

In constraints (5.24), the *spinning reserve* provided by on-line thermal power plants is represented by the term $\sum_{k \in K_z} (\overline{Q}_k^K \; \gamma_{k,t} - Q_{k,t}^K)$ and the spinning reserve provided by batteries is represented by the term $df \left[rt^D \; \overline{Q}^B \; (n_{z,0}^B + n_z^B) \; (1 - \phi_{z,t}) - B_{z,t}^D \right]$. The reserve margins are given by the quantity $rt^D \; \overline{Q}^B \; (n_{z,0}^B + n_z^B)$ the battery can discharge minus the quantity $B_{z,t}^D$ it is already discharging. If the battery is charging, i.e. $\phi_{z,t} = 1$ for the Constraint (5.15), then it cannot provide reserve. The battery can stop charging in order to meet the reserve requirement: the model can set $\phi_{z,t} = 0$ and $B_{z,t}^D = 0$ so that all the possible discharge quantities are available to provide reserve. The derating factor df accounts for battery degradation over time. Zonal reserve requirements can only be met by thermal generators and batteries located in the zone so that the reserve is not constrained by transmission. The slack variable[2] $RNP_{z,t}$ represents the Reserve Not Provided (RNP) and it is penalized in the objective function with the penalty coefficient c^{RNP} that should reflect the price of this RNP. In fact, the reserve required, $R_{z,t}$, is estimated using probabilistic approaches and depends on how much it costs and whether the requirement can be actually met.

If reserve margins can only be provided by on-line thermal power plants, the spinning reserve is the amount of unused capacity of on-line thermal power plants: in this case constraints (5.24) are replaced by constraints (5.43) for each $z \in Z$ and $t \in T$

$$\sum_{k \in K_z} (\overline{Q}_k^K \; \gamma_{k,t} - Q_{k,t}^K) + RNP_{z,t} \geq R_{z,t} \tag{5.43}$$

As an alternative, or in addition, to the spinning reserve, the so-called *substitution reserve* can be considered, i.e. the reserve provided by all the installed thermal power plants as in (5.44):

$$\sum_{k \in K_z} \left[\overline{Q}_k^K \; (n_{k,0}^K + n_k^K) - Q_{k,t}^K\right] + RNP_{z,t} \geq R_{z,t} \qquad z \in Z, t \in T \tag{5.44}$$

5.1.6 Price indeterminacy

When unique, the clearing price coincides with either an offer price or 0 or the VoLL. The zero value is associated with RES curtailment, which can never occur when considering a positive offer price and the possible partial acceptance of offer quantities, as explained in Section (4.1.4). The price can also assume a null value in particular cases, which are, however, the result of strategic, unrealistic behavior. The value of VoLL is reached only in case of load shedding (positive ENP). Since we are modeling an adequate system, there can be no load shedding and the price can never reach the VoLL, as explained in Section (4.1.6). Therefore, we exclude the possibility that the clearing price is 0 or the VoLL and focus only on it being equal to one of the offer prices. This is true when the clearing price is unique. In the case of indeterminacy, the clearing price can instead be unbounded from above, as shown in Fig. 4.6, being off-line generators treated as if they did not exist.

The CP may act strategically to determine these specific cases of indeterminacy and selecting, among the prices for which the MO is indifferent, the one that allows to meet the Revenue Adequacy constraints (see Section (5.1.7) later on), according to the optimistic view described

[2]With a non-standard definition, we might call it “pseudo” slack, being the constraint written as *available reserve* + $RNP \geq R$: the constraint is satisfied at equality when *available reserve* $\leq R$, with $RNP = R -$ *available reserve* or at inequality when *available reserve* $\geq R$, with $RNP = 0$.

in Section (3.2). To avoid this behavior, we formulate the innovative set of constraints (5.25) to (5.30), considering the following parameters

$\hat{\lambda}_i$	[€/MWh]	Offer price $i \in I$
$\hat{\lambda}_q$	[€/MWh]	Offer price $q \in Q$ of thermal power plants

The parameters $\hat{\lambda}_i$, for each i, represent the offer prices of both RES and thermal power plants, sorted in the ascending order $\{P^V, P^{WI}, P^{WO}, \hat{\lambda}_q\}$ for each q. The parameter $\hat{\lambda}_q$ specifically denotes the q-th offer price of thermal power plants of type $k \in K_q$: all the thermal generators belonging to the set K_q offer at the same price $\hat{\lambda}_q$.

A binary variable $y_{z,t,i}$ is introduced for each $\hat{\lambda}_i$. RES generators can only set the clearing price when dispatched: constraints (5.25) to (5.27) force the binary variables $y_{z,t,1}$, $y_{z,t,2}$ and $y_{z,t,3}$ to 0 if solar, inshore and offshore wind (respectively) are not dispatched in any of the zones. On the other hand, they can take the values 0 or 1 if the dispatched quantities are greater than 1 MWh, which is the minimum accepted quantity according to the market rules. The constraints take into account the sum of the dispatched quantities across the zones as, in the absence of congestion, the clearing price of a zone can be set by a generator located in any zone.

Similarly, only the on-line thermal generators can define the clearing price: by constraints (5.28) the binary variable $y_{z,t,3+q}$ is bound to 0 if no generator with offer price $\hat{\lambda}_q$ is on-line, while it can take the values 0 or 1 if at least one generator with offer price $\hat{\lambda}_q$ is on-line.

The clearing price $\lambda_{z,t}$ is a LL decision variable, and it can assume different values in the case of indeterminacy. In the UL we force to make it assume a single value as defined in the linear combination of the offer prices $\hat{\lambda}_i$, times the binaries $y_{z,t,i}$ working as activation variables, (5.30). As the clearing price can only be defined by one of the offer prices of the generators, then these variables must add up to 1, (5.29).

Because of constraints (5.29)-(5.30) any strategic attempt to create an indeterminate situation would be pointless, as they force the CP to choose the leftmost value of the indeterminacy interval. This selection of the left extreme can represent the CP acting as the regulator.

Note that the choice of the left extreme of the price indeterminacy interval is not guaranteed in the presence of generators dispatched at the technical minimum. For example, let us consider two generators, denoting by A the one offering at price a, and by B the one dispatched at the technical minimum offering at price $b > a$. Price indeterminacy occurs when A is dispatched at full capacity: the price is defined in the interval $[a, b]$. Since both a and b are considered by $\hat{\lambda}_i$, the chosen market clearing price could be equal to b.

A tighter version of constraints (5.29) and (5.30) is discussed in Section 5.3.

5.1.7 Revenue adequacy

The Revenue Adequacy constraints (5.31) guarantee that, for each type k of thermal power plant in each zone z, the sum of revenues from the market (from energy sales) $\sum_t \lambda_{z,t}\, Q^K_{k,t}$ and revenues from side payment (outside the market) u^K_k is at least equal to the sum of investment costs and operating costs, multiplied by the user-defined rate of return $M \geq 1$ expected by the plant owners. Focusing on the relevant role of the non-negative variable u^K_k, we note that such variable enters the RA constraint (5.31) as a slack variable[3]. If for thermal power plants of type k the market revenues exceed the total costs multiplied by M, the slack assumes a null value, and the constraint is satisfied with the strict inequality. Vice versa, if the market revenues are less than the total costs multiplied by M, the model determines the positive side payment u^K_k required to satisfy the RA constraint of plant type k at equality. In fact, the total side payments $\sum_{z \in Z} \sum_{k \in K_z} u^K_k$ appear in the OF as a term to be minimized.

In the light of what has been described in Section (1.3) about the Italian approach to the CRM, the values of u^K_k in the optimal solution give an indication of how much the specific technology needs to be compensated via side payments. This of course if the policymakers and regulators decide to adopt such measures. Additionally, choosing to satisfy the RA via discriminatory side payments, allows the CP not to increase the marginal price for all the energy produced and accepted by the MO.

We incidentally note that the model now becomes a mixed integer quadratically constrained due to the presence of the bilinear terms $\lambda_{z,t}\, Q^K_{k,t}$.

The term $\frac{|T|}{8760}$ makes investment and operating costs comparable as the former refer to the year while the latter to the representative days.

5.1.8 Objective function

Given the investment costs per MW

I^V_z	Solar investment cost in zone z
I^{WI}_z	Wind inshore investment cost in zone z
I^{WO}_z	Wind offshore investment cost in zone z
I^B_z	Battery investment cost in zone z
I^K_k	Investment cost of thermal power plants of cluster k

[3] With a non-standard definition, we might call it "pseudo" slack, being the constraint written as $revenues + u \geq costs$.

and the operational costs

C_k^M	Marginal costs of thermal power plants of cluster k
C_k^{SU}	Startup costs of thermal power plants of cluster k

the objective function (5.3), to be minimized, is the sum of four terms representing the investment costs, the operating costs of the thermal power plants, the side payments of the thermal power plants, and the penalties for reserve not provided, respectively. As the magnitudes of the last three terms depend on the number of days used to represent the target year, the investment cost is scaled to make it comparable with them, i.e. in the first term the annual investment cost is multiplied by $\frac{|T|}{8760}$. In the second term, the operating costs of the thermal power plants are the sum of startup costs and production costs.

Marginal costs and startup costs are considered for each thermal generator. The operating costs of the RES generators are not considered since they are null (in the LL problem, $P^V = LCOE$ represents the offer price while in the UL problem, only the investment costs are considered). Consequently, the UL problem can have alternative optimal solutions that differ in the value of the discharge $B_{z,t}^D$; these alternative optimal solutions for UL correspond to different LL costs.

5.2 GEP-RA (2)

In the GEP-RA (2) model, the CP determines the optimal system configuration by imposing the recovery of all investment costs (RES, batteries, thermal) and operating costs (thermal) through revenue adequacy constraints. In particular, this requires modeling the revenues of batteries.

The first and second terms of the GEP-RA (1) OF (5.3) do not appear in the GEP-RA (2) OF and are replaced by the following two terms:

- the total cost to consumers
- the side payments of power plants and batteries.

Model GEP-RA (2) allows to determine the best *trade-off* between non-discriminatory and discriminatory payments. The former are determined by the clearing price at which all electricity is sold, the latter are the different side payments specific for each plant or battery.

The GEP-RA (2) model is (5.45)-(5.61).

Objective Function

$$\min_{Var} \sum_{z\in Z}\sum_{t\in T} \lambda_{z,t} \left(D_{z,t} - B^D_{z,t} + B^C_{z,t}\right) + \sum_{z\in Z}\left(u^V_z + u^{WI}_z + u^{WO}_z + u^B_z\right) + \sum_{k\in K} u^K_k + \\ + \sum_{z\in Z}\sum_{t\in T} c^{RNP}\ RNP_{z,t} \tag{5.45}$$

Variables

$Var \equiv$

$\{(5.2)\}$ (GEP-RA (1)) $\cup$

$$\left\{u^V_z, u^{WI}_z, u^{WO}_z, u^B_z\right\} \text{(Side payments)} \tag{5.46}$$

Constraints on zonal capacities of RES power plants and batteries in the target year: $z \in Z$

$$(5.4) - (5.6) \tag{5.47}$$

Constraints on thermal power capacity in the target year: $k \in K_z,\ z \in Z$

$$(5.7) \tag{5.48}$$

UC constraints, with avoidance of physical withholding

$$(5.8) - (5.13) \tag{5.49}$$

Constraints that define accepted quantities and zonal prices for each hour: $t \in T$

$$\text{Optimality conditions of the MO problem (5.62)} \tag{5.50}$$

Constraints on operation of batteries: $z \in Z$

$$(5.15) - (5.16) \tag{5.51}$$

$$(5.18) - (5.23) \tag{5.52}$$

Reserve constraints: $z \in Z, t \in T$

$$(5.24) \tag{5.53}$$

Constraints that avoid price indeterminacy: $z \in Z,\ t \in T$

$$(5.25) - (5.30) \tag{5.54}$$

Revenue adequacy constraints: $z \in Z$

$$(5.31) \qquad k \in K_z \tag{5.55}$$

$$\sum_{t \in T} \lambda_{z,t}\ Q^V_{z,t} + u^V_z \geq M \frac{|T|}{8760}\ I^V_z\ \overline{Q}^V\ n^V_z \tag{5.56}$$

$$\sum_{t \in T} \lambda_{z,t}\ Q^{WI}_{z,t} + u^{WI}_z \geq M \frac{|T|}{8760}\ I^{WI}_z\ \overline{Q}^{WI}\ n^{WI}_z \tag{5.57}$$

$$\sum_{t \in T} \lambda_{z,t}\ Q^{WO}_{z,t} + u^{WO}_z \geq M \frac{|T|}{8760}\ I^{WO}_z\ \overline{Q}^{WO}\ n^{WO}_z \tag{5.58}$$

$$\sum_{t \in T} \lambda_{z,t}\ (B^D_{z,t} - B^C_{z,t}) + u^B_z \geq M \frac{|T|}{8760}\ I^B_z\ \overline{Q}^B\ n^B_z \tag{5.59}$$

Variables types

$$(5.32) - (5.39) \tag{5.60}$$

$$u^V_z, u^{WI}_z, u^{WO}_z, u^B_z \geq 0 \qquad z \in Z \tag{5.61}$$

5.2.1 Determination of accepted quantities and zonal prices

To introduce RA constraints for batteries (see Section (5.2.2)), we need to express their market revenues. This in turn requires the modified model (5.62) for the LL.

$$\min_{Q, F_{l,t}} (5.42a) \tag{5.62a}$$

$$Q^V_{z,t} + Q^{WI}_{z,t} + Q^{WO}_{z,t} + \sum_{k \in K_z} Q^K_{k,t} + B^D_{z,t}$$

$$+ \sum_{l \in BS_z} F_{l,t} \geq D_{z,t} + \sum_{l \in FS_z} F_{l,t} + B^C_{z,t} \qquad z \in Z \qquad (\lambda_{z,t} \geq 0) \tag{5.62b}$$

$$(5.42c) - (5.42g) \tag{5.62c}$$

Balance constraints (5.62b) consider discharge as an increase in supply and charge as an increase in demand. The clearing price $\lambda_{z,t}$ is determined by the MO, taking into account the participation of the batteries in the system operation (the batteries do not offer, but influence the clearing price through their discharge-charge dynamics).

Note that the increase of the load determined by the charge can be met by both RES and thermal productions. The UL constraint (5.17), limiting the charge to the RES production net of the quantity accepted by the MO, is no longer considered. RES now also have revenues for battery charging, which in this model is accepted on the market.

5.2.2 Revenue Adequacy

In addition to the RA constraints (5.31) relating to thermal power plants, the additional RA constraints (5.56)-(5.58) relate to RES, and constraints (5.59) relate to batteries. Note that investment costs only appear on the right-hand side of these new RA constraints, as operating costs are zero for the respective technologies.

Similarly to the role of side payments for thermal plants, here the slack variables u^V_z, u^{WI}_z and u^{WO}_z represent the side payments for RES technologies. In line with the spirit of the consultation [Ministero dell'Ambiente e della Sicurezza Energetica, 2023], they can be interpreted as the amount of economic resources that an auction can give. Such discriminatory remunerations are used by the system to support RES that otherwise may fail to stay in the market but are needed to achieve policy objectives. At the same time, these auctions tend to stabilize energy market prices by imposing two-ways CfD (which we do not model and can be computed off-line).

In particular, in constraints (5.59), the first term represents the net revenue of the batteries, i.e. the difference between what is received when discharging and what is paid when charging. We recall that by constraints (5.15), (5.16) charge and discharge are mutually exclusive in any hour. The second term represents the side payments possible needed to cover the investment costs. Depending on the design, they could be seen as either discriminatory compensation to batteries from the capacity market or additional revenue from the re-dispatch markets.

5.2.3 Objective function

The objective function (5.45), to be minimized, represents the total cost for consumers. The first term represents the costs paid on the market, while the second and third terms express the side payments, which are ultimately also paid by consumers as they are included in their electricity bills.

The first term represents the value of the produced energy, considering for each zone z and hour t either the energy $(D_{z,t} + B^C_{z,t})$ if the battery is charging or the energy $(D_{z,t} - B^D_{z,t})$ if the battery is discharging. In this way, the energy actually produced in each zone z and hour t is evaluated at the corresponding market clearing price $\lambda_{z,t}$.

Unlike the GEP-RA (1) OF (5.3), the GEP-RA (2) OF (5.45) allows the CP to decide the operation of batteries taking into account the impact of such decision on the market clearing price determined by the LL. In fact, different values of battery discharge may correspond to different accepted quantities, and thus to different LL costs and clearing prices. Therefore, since clearing prices appear in the GEP-RA (2) OF, the CP optimizes the use of the batteries.

In this version, new bilinear terms are introduced both in the OF ($B^D_{z,t}\ \lambda_{z,t}$, $B^C_{z,t}\ \lambda_{z,t}$) and in the RA constraints ($B^D_{z,t}\ \lambda_{z,t}$, $B^C_{z,t}\ \lambda_{z,t}$, $\lambda_{z,t}\ Q^V_{z,t}$, $\lambda_{z,t}\ Q^{WI}_{z,t}$, $\lambda_{z,t}\ Q^{WO}_{z,t}$).

5.3 An improved version of the constraints on the choice of the left extreme of the price indeterminacy interval

The use of binary variables $y_{z,t,i}$ with the requirement of constraint (5.29) that they add up to one can slow down branch-and-bound MIP solvers [Yıldız and Vielma, 2013]. Specifically, this constraint leads to unbalanced branch-and-bound trees when $|I| > 2$. Indeed, when variable $y_{z,t,i}$ is set to 1 (*up-branching*), all other variables $y_{z,t,j}$ for $j \in I - \{i\}$ are set to 0; on the contrary, when $y_{z,t,i}$ is set to 0 (*down-branching*), the other variables are let unconstrained, which may slow down the solution process. To address this balancing issue, [Vielma, 2015] proposes an *extended formulation*, which we introduce in our model by defining the auxiliary binary variables $w_{z,t,i}$ and replacing constraints (5.29) and (5.30) by (5.63)-(5.65).

$$\lambda_{z,t} = \hat{\lambda}_1 w_{z,t,1} + \sum_{i=2}^{|I|} (\hat{\lambda}_i - \hat{\lambda}_{i-1}) w_{z,t,i} \qquad z \in Z, t \in T \quad (5.63)$$

$$w_{z,t,i} \geq w_{z,t,i+1} \qquad i \in [1 : |I| - 1], z \in Z, t \in T \quad (5.64)$$

$$w_{z,t,1} = 1 \qquad z \in Z, t \in T \quad (5.65)$$

If $w_{z,t,1} = 1$ and $w_{z,t,2} = ... = w_{z,t,|I|} = 0$, then $\lambda_{z,t} = \hat{\lambda}_1$; if $w_{z,t,1} = w_{z,t,2} = 1$ and $w_{z,t,3} = ... = w_{z,t,|I|} = 0$, then $\lambda_{z,t} = \hat{\lambda}_1 + \hat{\lambda}_2 - \hat{\lambda}_1 = \hat{\lambda}_2$ and so on. For $i \geq 2$, constraint (5.64) implies that constraint (5.63) becomes $\lambda_{z,t} = \hat{\lambda}_i$ since for $1 \leq j \leq i - 1$, $\hat{\lambda}_j$ and $-\hat{\lambda}_j$ cancel out. Therefore, as in the original formulation, only one price is selected among the offers. However, the branching is more efficient in the extended formulation: indeed, for $i = 1 : |I| - 1$, $w_{z,t,i} = 0$ (down-branching) implies $w_{z,t,j} = 0$ for $j > i$, while for $i = 2 : |I|$, $w_{z,t,i} = 1$ (up-branching) implies $w_{z,t,j} = 1$ for $j < i$. This effect is named double contracting branching.

To preserve the validity of the constraints (5.25)-(5.28) that depend on the variable $y_{z,t,i}$, we introduce (5.66) and (5.67) to link $y_{z,t,i}$ to the auxiliary variable $w_{z,t,i}$.

$$y_{z,t,i} = w_{z,t,i} - w_{z,t,i+1} \qquad i \in [1 : |I| - 1], z \in Z, t \in T \quad (5.66)$$

$$y_{z,t,i} = w_{z,t,i} \qquad i = |I|, z \in Z, t \in T \quad (5.67)$$

In fact, the binary $y_{z,t,i}$ can take the value 1, with a corresponding clearing price $\lambda_{z,t} = \hat{\lambda}_i$, only if $w_{z,t,i} = 1$ and $w_{z,t,i+1} = 0$.

5.4 Mixed-binary version

We also created a **mixed-binary version** of the model to test whether a computational improvement can be achieved. This model has been obtained by applying a *binary expansion* to each integer variable considered in the main model. In particular, for thermal power plants, we replace the integer variables n_k^K, $\gamma_{k,t}$, $\alpha_{k,t}$ and $\beta_{k,t}$, with their binary expansion, determined by introducing the binary variables

- $n_{k,e}^{K-BIN} \in \{0, 1\}$, with $e \in E_k^K$, which is the set of exponents of 2 needed to express the maximum number of new generators of cluster k ($\overline{n}_k^K$)

- $\gamma_{k,t,e}^{BIN}$, $\alpha_{k,t,e}^{BIN}$, $\beta_{k,t,e}^{BIN} \in \{0, 1\}$, with $e \in E_k^{K+K0}$, which is the set of exponents of 2 needed to express the sum of the maximum number of new generators ($\overline{n}_k^K$) and the number of existing generators (n_k^{K0}).

For example, if $\overline{n}_k^K = 7$ and $n_k^{K0} + \overline{n}_k^K = 9$, then $E_k^K = \{0,1,2\}$ and $E_k^{K+K0} = \{0,1,2,3\}$. The variable n_k^K is replaced by

$$\sum_{e \in E_k^K} 2^e \; n_{k,e}^{K-BIN} = 2^0 \; n_{k,0}^{BIN} + 2^1 \; n_{k,1}^{BIN} + 2^2 \; n_{k,2}^{BIN}$$

while variable $\gamma_{k,t}$ is replaced by

$$\sum_{e \in E_k^{K+K0}} 2^e \; \gamma_{k,t,e}^{BIN} = 2^0 \; \gamma_{k,t,0}^{BIN} + 2^1 \; \gamma_{k,t,1}^{BIN} + 2^2 \; \gamma_{k,t,2}^{BIN} + 2^3 \gamma_{k,t,3}^{BIN}$$

Constraints (5.7) and (5.40) are replaced by (5.68) and (5.69).

$$\sum_{e \in E_k^K} n_{k,e}^{K-BIN} \leq \overline{n}_k^K \qquad k \in K \qquad (5.68)$$

$$\sum_{e \in E_k^{K+K0}} 2^e \; \gamma_{k,t,e}^{BIN} \leq n_{k,0}^K + \sum_{e \in E_k^K} n_{k,e}^{K-BIN} \qquad k \in K, t \in T \quad (5.69)$$

For RES generators and batteries, we replace the integer variables $n_z^V, n_z^{WI}, n_z^{WO}, n_z^B$ with their binary expansion, determined by introducing the binary variables

- $n_{z,e}^{V-BIN} \in \{0,1\}$, with $e \in E_z^V$, which is the set of exponents of 2 needed to express the maximum number of new solar generators in zone z
- $n_{z,e}^{WI-BIN} \in \{0,1\}$, with $e \in E_z^{WI}$, which is the set of exponents of 2 needed to express the maximum number of new wind inshore generators in zone z
- $n_{z,e}^{WO-BIN} \in \{0,1\}$, with $e \in E_z^{WO}$, which is the set of exponents of 2 needed to express the maximum number of new wind offshore generators in zone z
- $n_{z,e}^{B-BIN} \in \{0,1\}$, with $e \in E_z^B$, which is the set of exponents of 2 needed to express the maximum number of new batteries in zone z

Constraints (5.4) and (5.5) are replaced by (5.70) and (5.71)

$$\underline{V}_z \leq \overline{Q}^V (n_{z,0}^V + \sum_{e \in E_z^V} n_{z,e}^{V-BIN}) \leq \overline{V}_z \qquad z \in Z \quad (5.70)$$

$$\underline{W}_z \leq \overline{Q}^{WI} (n_{z,0}^{WI} + \sum_{e \in E_z^{WI}} n_{z,e}^{WI-BIN}) + \overline{Q}^{WO} (n_{z,0}^{WO} + \sum_{e \in E_z^{WO}} n_{z,e}^{WO-BIN}) \leq \overline{W}_z \qquad z \in Z \quad (5.71)$$

Note that different E sets were created to define the right number of binaries to represent the maximum value that different integer variables can take and not a larger number of binaries when these are not needed.

By replacing all the integer variables with their binary expansions, the GEP-RA (1) and GEP-RA (2) models are redefined, resulting in two models consisting solely of continuous and binary variables.

Chapter 6

Resolution of the proposed models

In Section (6.1), we present various reformulations of the lower-level problems discussed earlier. Certain reformulations incorporate bilinear terms, and these, along with the bilinear products in the RA constraints and objective function, are linearized in Section (6.2).

6.1 Reformulation of the bilevel model into a single-level model

The LL problems (5.42) and (5.62), which we refer to as LL (1) and LL (2) respectively, are linear and can be included in the respective UL problem as a set of linear and non-linear equations obtained by applying the *Karush-Kuhn-Tucker optimality conditions* (KKTs), as described in Chapter 3. In this way, the bilevel model is transformed into a single-level model. The KKTs are determined as the combination of the primal constraints, dual constraints, and the Complementary Slackness Conditions (CSC), which require a null product of each primal constraint with its corresponding dual variable (both in the form ≥ 0) and vice versa.

$0 \leq cp \perp vd \geq 0$

$0 \leq cd \perp vp \geq 0$

Here, cp denotes each primal constraint, vd denotes the corresponding dual variable, cd denotes each dual constraint, and vp denotes the corresponding primal variable. In Appendix A is the explicit formulation of both the primal-dual constraints and the different reformulations for CSC, which are reported here in compact form.

The CSC of LL (2) in the *Product* formulation are defined as (6.1) for each $t \in T$.

$$\left(Q^V_{z,t}+Q^{WI}_{z,t}+Q^{WO}_{z,t}+\sum_{k\in K_z} Q^K_{k,t}+\sum_{l\in BS_z} F_{l,t}-\sum_{l\in FS_z} F_{l,t}-D_{z,t}+B^D_{z,t}-B^C_{z,t}\right) \perp \lambda_{z,t} \quad z \in Z \quad (6.1a)$$

$$(F_{l,t} - \underline{F}_l) \perp \underline{\upsilon}_{l,t} \quad l \in L \quad (6.1b)$$

$$\left(\overline{F}_l - F_{l,t}\right) \perp \overline{\upsilon}_{l,t} \quad l \in L \quad (6.1c)$$

$$\left(cf^V_{z,t} \cdot \overline{Q}^V \cdot (n^V_{z,0} + n^V_z) - Q^V_{z,t}\right) \perp \mu^V_{z,t} \quad z \in Z \quad (6.1d)$$

$$\left(cf^{WI}_{z,t} \cdot \overline{Q}^{WI} \cdot (n^{WI}_{z,0} + n^{WI}_z) - Q^{WI}_{z,t}\right) \perp \mu^{WI}_{z,t} \quad z \in Z \quad (6.1e)$$

$$\left(cf^{WO}_{z,t} \cdot \overline{Q}^{WO} \cdot (n^{WO}_{z,0} + n^{WO}_z) - Q^{WO}_{z,t}\right) \perp \mu^{WO}_{z,t} \quad z \in Z \quad (6.1f)$$

$$\left(Q^K_{k,t} - \underline{Q}^K_k \cdot \gamma_{k,t}\right) \perp \underline{\mu}^K_{k,t} \quad k \in K \quad (6.1g)$$

$$\left(\overline{Q}^K_k \cdot \gamma_{k,t} - Q^K_{k,t}\right) \perp \overline{\mu}^K_{k,t} \quad k \in K \quad (6.1h)$$

$$\left(P^V - \lambda_{z,t} + \mu^V_{z,t}\right) \perp Q^V_{z,t} \quad z \in Z \quad (6.1i)$$

$$\left(P^{WI} - \lambda_{z,t} + \mu^{WI}_{z,t}\right) \perp Q^{WI}_{z,t} \quad z \in Z \quad (6.1j)$$

$$\left(P^{WO} - \lambda_{z,t} + \mu^{WO}_{z,t}\right) \perp Q^{WO}_{z,t} \quad z \in Z \quad (6.1k)$$

$$\left(P^K_k - \lambda_{z,t} - \underline{\mu}^K_{k,t} + \overline{\mu}^K_{k,t}\right) \perp Q^K_{k,t} \quad k \in K_z, z \in Z \quad (6.1l)$$

For the LL (1) model, the CSC (6.1a) does not include the term $-B^C_{z,t}$ in the left-hand term (primal constraint).

Note that when the CSC are included in the UL as additional constraints, the terms

$$B^C_{z,t}, B^D_{z,t}, n^V_z, n^{WI}_z, n^{WO}_z, \gamma_{k,t}$$

are no longer parameters but represent variables, determining new bilinear terms.

The nonlinear conditions (6.1a)-(6.1l) can be reformulated with different methods; depending on the chosen reformulation, bilinear terms may or may not be considered. We tested different methods by both deriving the reformulations manually and using the automatic reformulations available in the BilevelJuMP package [Dias Garcia et al., 2023] belonging to the Julia Programming Language ecosystem. Concerning the latter, the technology aspects will be detailed in (8.1).

The first reformulation, see (A.3), consists of aggregating all the conditions into a *single* *constraint* in which all the terms (written in the $\geq$ form) sum up to 0:

$$cp \cdot vd + cd \cdot vp = 0 \tag{6.2}$$

The *Special Order Sets of type 1* (SOS1) reformulation, see (A.4), consists of introducing a set of variables where at most one variable in the set can take a value greater than zero. Given $\{s_1^A, s_2^A\} \in SOS1$ and $\{s_1^B, s_2^B\} \in SOS1$, the CSC can be rewritten as follows

$$s_1^A = cp \qquad s_2^A = vd$$
$$s_1^B = cd \qquad s_2^B = vp$$

The *Fortuny-Amat and McCarl* (Big-M) [Fortuny-Amat and McCarl, 1981] reformulation, see (A.5), consists of introducing a binary for each complementarity condition. Given $b1 \in \{0, 1\}$ and $b2 \in \{0, 1\}$ the CSC can be rewritten as follows

$$cp \leq M^{cp}\ (1 - b1) \qquad vd \leq M^{vd}\ b1$$
$$cd \leq M^{cd}\ (1 - b2) \qquad vp \leq M^{vp}\ b2$$

where M^{cp}, M^{vd}, M^{cd}, M^{vp} are large enough numbers so that the optimal solution(s) of the problem is not cut off. If $b1 = 1$, then $cp = 0$ while the constraint on the associated dual variable vd is redundant; if $b1 = 0$, then $vd = 0$ and the constraint on cp is redundant. The complete formulation is reported in Appendix

To find reasonable bounds, we evaluate each term of the CSC, as shown in Tab. 6.1 and Tab. 6.2. Note that bounds on the primal variables are easy to derive, as they are bounded by the physics of the problem, modeled in the UL constraints. For the dual variables, we base our bounds on the interpretation of the dual of the LL. In particular, when $cf_{z,t}^V = 0$, there is no upper bound on the dual variables $\mu_{z,t}^V$ (and the same for those related to wind), that can take any value $\geq \lambda_{z,t}^* - P^V$ without affecting the optimal solution. Therefore, we consider the upper bound $\overline{\lambda} - P^V$ for both $cf_{z,t}^V > 0$ and $cf_{z,t}^V = 0$, as it is the real upper bound in the first case, while it is among the values that do not affect the objective function in the second case. Similarly for wind variables.

When $\gamma_{k,t}^* = 0$, there is no upper bound on the dual variables $\underline{\mu}_{k,t}^K$ and $\overline{\mu}_{k,t}^K$, that can take any value such that $\overline{\mu}_{k,t}^K - \underline{\mu}_{k,t}^K \geq \lambda_{z,t}^* - P_k^K$ without affecting the optimal solution. For $\overline{\mu}_{k,t}^K$ and $\underline{\mu}_{k,t}^K$ we respectively consider the upper bounds $\overline{\lambda} - P_k^K$ and P_k^K as they are the real upper bounds for $\gamma_{k,t}^* > 0$, while they are among the values that do not affect the objective function

cp	M^{cp}	vd	M^{vd}		from
			$cf, \gamma^* > 0$	$cf, \gamma^* = 0$	
$Q^V_{z,t} + Q^{WI}_{z,t} + Q^{WO}_{z,t} + \sum_{k\in K_z} Q^K_{k,t} + \sum_{l\in BS_z} F_{l,t} - \sum_{l\in FS_z} F_{l,t} - D_{z,t} + B^D_{z,t} - B^C_{z,t}$	0	$\lambda_{z,t}$	$\overline{\lambda} = \max_{i\in I} \hat{\lambda}_i$		(4.24)
$F_{l,t} - \underline{F}_l$	$\overline{F}_l - \underline{F}_l$	$\underline{\upsilon}_{l,t}$	$\overline{\lambda}$		(4.9)
$\overline{F}_l - F_{l,t}$	$\overline{F}_l - \underline{F}_l$	$\overline{\upsilon}_{l,t}$	$\overline{\lambda}$		(4.9)
$cf^V_{z,t}\overline{Q}^V(n^V_{z,0} + n^V_z) - Q^V_{z,t}$	$cf^V_{z,t}\overline{V}_z$	$\mu^V_{z,t}$	$\overline{\lambda} - P^V$	∞	(4.19)
$cf^{WI}_{z,t}\overline{Q}^{WI}(n^{WI}_{z,0} + n^{WI}_z) - Q^{WI}_{z,t}$	$cf^{WI}_{z,t}\overline{W}_z$	$\mu^{WI}_{z,t}$	$\overline{\lambda} - P^{WI}$	∞	(4.19)
$cf^{WO}_{z,t}\overline{Q}^{WO}(n^{WO}_{z,0} + n^{WO}_z) - Q^{WO}_{z,t}$	$cf^{WO}_{z,t}\overline{W}_z$	$\mu^{WO}_{z,t}$	$\overline{\lambda} - P^{WO}$	∞	(4.19)
$Q^K_{k,t} - \underline{Q}^K_k \gamma_{k,t}$	$\overline{Q}^K_k(n^K_{k,0} + \overline{n}^K_k)$	$\underline{\mu}^K_{k,t}$	$P^K_k(\star)$	∞	(4.13)
$\overline{Q}^K_k \gamma_{k,t} - Q^K_{k,t}$	$\overline{Q}^K_k(n^K_{k,0} + \overline{n}^K_k)$	$\overline{\mu}^K_{k,t}$	$\overline{\lambda} - P^K_k$	∞	(4.13)

Table 6.1: Upperbounds M^{cp} of primal constraints cp and M^{vd} of dual variables vd. $(\star)$ P^K_k instead of $P^K_k - \overline{\lambda}$ as the latter could be negative.

for $\gamma^*_{k,t} = 0$.

cd	M^{cd}	vp	M^{vp}	from
$P^V - \lambda_{z,t} + \mu^V_{z,t}$	$P^V + \overline{\lambda} - P^V$	$Q^V_{z,t}$	$cf^V_{z,t}\overline{V}_z$	(5.42c), (5.4)
$P^{WI} - \lambda_{z,t} + \mu^{WI}_{z,t}$	$P^{WI} + \overline{\lambda} - P^{WI}$	$Q^{WI}_{z,t}$	$cf^{WI}_{z,t}\overline{W}_z$	(5.42d), (5.5)
$P^{WO} - \lambda_{z,t} + \mu^{WO}_{z,t}$	$P^{WO} + \overline{\lambda} - P^{WO}$	$Q^{WO}_{z,t}$	$cf^{WO}_{z,t}\overline{W}_z$	(5.42e), (5.5)
$P^K_k - \lambda_{z,t} - \underline{\mu}^K_{k,t} + \overline{\mu}^K_{k,t}$	$P^K_k + \overline{\lambda} - P^K_k$	$Q^K_{k,t}$	$\overline{Q}^K_k(n^K_{k,0} + \overline{n}^K_k)$	(5.42f), (5.10), (5.7)

Table 6.2: Upper bounds M^{cd} of dual constraints cd and M^{vp} of primal variables vp

It is worth noting that the upper bounds on variables of Tab. 6.1 and Tab. 6.2 are useful not only for the Big-M reformulation of the KKTs, but more generally in helping the solver to define the feasible region. Indeed, aggressive bound tightening represents an important building block for enhancing MIP technology in the presence of indicator constraints and disjunctive terms, as described in [Belotti et al., 2016]. Therefore, we introduce upper bound constraints in all the formulations.

The *Strong Duality* (SD) reformulation, see (A.6), consists in replacing the CSC by a single constraint (either with the sum over $t \in T$ or for each $t \in T$) that requires the primal objective function LLP to be equal to the dual objective function LLD:

$$LLP = LLD \tag{6.3}$$

The manual formulation of the single-level model allows a two-step solution process: initially, a solution is derived by relaxing the CSC (see Def. (3.2.3)), and subsequently, this solution is employed as a warm start for the complete model.

6.2 Linearized version: an exact linearization scheme

Bilinear terms occur

- in the objective function (5.45) $B^C_{z,t}\ \lambda_{z,t}, B^D_{z,t}\ \lambda_{z,t}$
- in the RA constraints (5.31), (5.56)-(5.59) $Q^K_{k,t}\ \lambda_{z,t}$, $Q^V_{z,t}\ \lambda_{z,t}$, $Q^{WI}_{z,t}\ \lambda_{z,t}$, $Q^{WO}_{z,t}\ \lambda_{z,t}$, $B^C_{z,t}\ \lambda_{z,t}$, $B^D_{z,t}\ \lambda_{z,t}$
- (possibly) in the strong duality constraint $B^C_{z,t}\ \lambda_{z,t}$, $B^D_{z,t}\ \lambda_{z,t}$, $n^V_z\ \mu^V_{z,t}$, $n^{WI}_z\ \mu^{WI}_{z,t}$, $n^{WO}_z\ \mu^{WO}_{z,t}$, $\gamma_{k,t}\ \underline{\mu}^K_{k,t}$, $\gamma_{k,t}\ \overline{\mu}^K_{k,t}$

and can be handled directly by the Gurobi solver (since it can process MIQCP models) or reformulated (exactly) as follows to create MIP versions of the models.

Bilinear terms that are products between continuous variables ($B\ \lambda$, $Q\ \lambda$) are linearized through a *Binary Expansion*, followed by a *McCormick Relaxation*, [McCormick, 1976]. These are a generalization of the linearization proposed in the seminal work [Fortet, 1960] which considers products of only binary variables with the introduction of an extended variable and a set of linear constraints.

Indeed, for constraints (5.30),(5.29), the clearing price λ can be written as the linear combination $\lambda = \sum_{i\in I} \hat{\lambda}_i\ y_i$, where binaries y_i add up to 1. Therefore, we can replace λ with its linear combination

$$B\ \lambda = B\ \sum_{i\in I} \hat{\lambda}_i\ y_i$$

$$Q\ \lambda = Q\ \sum_{i\in I} \hat{\lambda}_i\ y_i$$

Since the new bilinear terms are products of bounded continuous and binary variables, the *McCormick Relaxation technique* can be applied without introducing any approximation. However, this technique is weak as it is exact only for $y_i = 1$ or $y_i = 0$ and not for any other value that the branch and bound finds in relaxing the binary.

Auxiliary continuous variables $z^B_i = B\ y_i$, $z^Q_i = Q\ y_i$ are introduced so that

$$B\ \lambda = B\ \sum_{i\in I} \hat{\lambda}_i\ y_i = \sum_{i\in I} \hat{\lambda}_i\ z^B_i$$

$$Q\ \lambda = Q\ \sum_{i\in I} \hat{\lambda}_i\ y_i = \sum_{i\in I} \hat{\lambda}_i\ z^Q_i$$

The following linear constraints link the variables

$$0 \leq B - z_i^B \leq M^B\ (1 - y_i)$$
$$0 \leq z_i^B \leq M^B\ y_i$$
$$0 \leq Q - z_i^Q \leq M^Q\ (1 - y_i)$$
$$0 \leq z_i^Q \leq M^Q\ y_i$$

Indeed, if $y_i = 0$, then $z_i^B = 0$, and the constraint $B \leq M^B$ is redundant (z_i^B and B are not related); if $y_i = 1$, then $z_i^B = B$. Similarly for z_i^Q and Q. The constants M^B and M^Q must be chosen large enough not to constrain the variables and cut off their optimal values, but also not too large to avoid weak LP relaxations. We considered the values shown in Tab. 6.3. Note

Variable	M	from
$B^D_{z,t}$	$D_{z,t}$	(5.16)
$B^C_{z,t}$	$rt^C\ \overline{B}_z$	(5.18), (5.6)
$Q^V_{z,t}, Q^{WI}_{z,t}, Q^{WO}_{z,t}, Q^K_{k,t}$	Tab. 6.2	

Table 6.3: Upper bounds M for the McCormick relaxation constraints

that while discharge is limited by demand, because it makes no sense for batteries to discharge except to cover demand, the charge could be higher. In fact, if the batteries have the capacity to charge and the demand is low, the remaining capacity of the generators can be used to charge the batteries. In the GEP-RA (1) model, a possibly tighter value could be defined, as the charge is limited by the production of RES according to the constraint (5.17), which in turn is limited by the constraints (5.4), (5.5) on RES maximum capacity: $cf^V_{z,t}\ \overline{V}_z + cf^{WI}_{z,t}\ \overline{W}_z + cf^{WO}_{z,t}\ \overline{W}_z$. If the user decides to make the constraints (5.4), (5.5) and (5.6) inactive, the Big-Ms are set to 10^6. In particular, the revenues (for each $z \in Z, t \in T$) are rewritten as

$$Q^K_{k,t}\ \lambda_{z,t} = Q^K_{k,t} \sum_{i \in I} \hat{\lambda}_i\ y_{z,t,i} = \sum_{i \in I} \hat{\lambda}_i\ z^K_{k,t,i} \qquad k \in K_z$$

$$Q^V_{z,t}\ \lambda_{z,t} = Q^V_{z,t} \sum_{i \in I} \hat{\lambda}_i\ y_{z,t,i} = \sum_{i \in I} \hat{\lambda}_i\ z^V_{z,t,i}$$

$$Q^{WI}_{z,t}\ \lambda_{z,t} = Q^{WI}_{z,t} \sum_{i \in I} \hat{\lambda}_i\ y_{z,t,i} = \sum_{i \in I} \hat{\lambda}_i\ z^{WI}_{z,t,i}$$

$$Q^{WO}_{z,t}\ \lambda_{z,t} = Q^{WO}_{z,t} \sum_{i \in I} \hat{\lambda}_i\ y_{z,t,i} = \sum_{i \in I} \hat{\lambda}_i\ z^{WO}_{z,t,i}$$

$$B^D_{z,t}\ \lambda_{z,t} - B^C_{z,t}\ \lambda_{z,t} = B^D_{z,t} \sum_{i \in I} \hat{\lambda}_i\ y_{z,t,i} - B^C_{z,t} \sum_{i \in I} \hat{\lambda}_i\ y_{z,t,i} = \sum_{i \in I} \hat{\lambda}_i\ z^{BD}_{z,t,i} - \sum_{i \in I} \hat{\lambda}_i\ z^{BC}_{z,t,i}$$

so that the RA constraints (5.31), (5.56) - (5.59) become, respectively

$$\sum_{t\in T}\sum_{i\in I}\hat{\lambda}_i\ z^K_{k,t,i} + u^K_k \geq M\left[\sum_{t\in T}(C^M_k\ Q^K_{k,t} + C^{SU}_k\ \alpha_{k,t}) + \frac{|T|}{8760}\ I^K_k\ \overline{Q}^K_k\ n^K_k\right] \qquad k\in K \quad (6.4)$$

$$\sum_{t\in T}\sum_{i\in I}\hat{\lambda}_i\ z^V_{z,t,i} + u^V_z \geq M\frac{|T|}{8760}\ I^V_z\ \overline{Q}^V\ n^V_z \qquad z\in Z \quad (6.5)$$

$$\sum_{t\in T}\sum_{i\in I}\hat{\lambda}_i\ z^{WI}_{z,t,i} + u^{WI}_z \geq M\frac{|T|}{8760}\ I^{WI}_z\ \overline{Q}^{WI}\ n^{WI}_z \qquad z\in Z \quad (6.6)$$

$$\sum_{t\in T}\sum_{i\in I}\hat{\lambda}_i\ z^{WO}_{z,t,i} + u^{WO}_z \geq M\frac{|T|}{8760}\ I^{WO}_z\ \overline{Q}^{WO}\ n^{WO}_z \qquad z\in Z \quad (6.7)$$

$$\sum_{t\in T}\sum_{i\in I}\hat{\lambda}_i\ z^{BD}_{z,t,i} - \sum_{t\in T}\sum_{i\in I}\hat{\lambda}_i\ z^{BC}_{z,t,i} + u^B_z \geq M\frac{|T|}{8760}\ I^B_z\ \overline{B}\ n^B_z \qquad z\in Z \quad (6.8)$$

To link the variables $z^K_{k,t,i}$, $y_{z,t,i}$, and $Q^K_{k,t}$, the following constraints are introduced:

$$0 \leq Q^K_{k,t} - z^K_{k,t,i} \leq \overline{Q}^K_k\ (n^K_{k,0} + \overline{n}^K_k)\ (1 - y_{z,t,i}) \qquad k\in K_z, i\in I, z\in Z, t\in T \quad (6.9)$$

$$0 \leq z^K_{k,t,i} \leq \overline{Q}^K_k\ (n^K_{k,0} + \overline{n}^K_k)\ y_{z,t,i} \qquad k\in K_z, i\in I, z\in Z, t\in T \quad (6.10)$$

Similarly for the other variables:

$$0 \leq Q^V_{z,t} - z^V_{z,t,i} \leq cf^V_{z,t}\ \overline{V}_z\ (1 - y_{z,t,i}) \qquad i\in I, z\in Z, t\in T \quad (6.11)$$

$$0 \leq z^V_{z,t,i} \leq cf^V_{z,t}\ \overline{V}_z\ y_{z,t,i} \qquad i\in I, z\in Z, t\in T \quad (6.12)$$

$$0 \leq Q^{WI}_{z,t} - z^{WI}_{z,t,i} \leq cf^{WI}_{z,t}\ \overline{W}_z\ (1 - y_{z,t,i}) \qquad i\in I, z\in Z, t\in T \quad (6.13)$$

$$0 \leq z^{WI}_{z,t,i} \leq cf^{WI}_{z,t}\ \overline{W}_z\ y_{z,t,i} \qquad i\in I, z\in Z, t\in T \quad (6.14)$$

$$0 \leq Q^{WO}_{z,t} - z^{WO}_{z,t,i} \leq cf^{WO}_{z,t}\ \overline{W}_z\ (1 - y_{z,t,i}) \qquad i\in I, z\in Z, t\in T \quad (6.15)$$

$$0 \leq z^{WO}_{z,t,i} \leq cf^{WO}_{z,t}\ \overline{W}_z\ y_{z,t,i} \qquad i\in I, z\in Z, t\in T \quad (6.16)$$

$$0 \leq B^D_{z,t} - z^{BD}_{z,t,i} \leq D_{z,t}\ (1 - y_{z,t,i}) \qquad i\in I, z\in Z, t\in T \quad (6.17)$$

$$0 \leq z^{BD}_{z,t,i} \leq D_{z,t}\ y_{z,t,i} \qquad i\in I, z\in Z, t\in T \quad (6.18)$$

$$0 \leq B^C_{z,t} - z^{BC}_{z,t,i} \leq rt^C\ \overline{B}_z\ (1 - y_{z,t,i}) \qquad i\in I, z\in Z, t\in T \quad (6.19)$$

$$0 \leq z^{BC}_{z,t,i} \leq rt^C\ \overline{B}_z\ y_{z,t,i} \qquad i\in I, z\in Z, t\in T \quad (6.20)$$

Bilinear terms that are products between integer and continuous variables ($n\ \mu$, $\gamma\ \underline{\mu}$, $\gamma\ \overline{\mu}$) are linearized through a *Binary Expansion* of the integer variables, as described in (5.4), followed by a *McCormick Relaxation*. Indeed, by introducing a set $E = \{0, ..., \overline{E}\}$ of exponent of 2, one can rewrite the integer variables n, γ as a combination of auxiliary binary variables n_e^{BIN}, γ_e^{BIN}:

$$n = \sum_{e \in E} 2^e \ n_e^{BIN}$$
$$\gamma = \sum_{e \in E} 2^e \ \gamma_e^{BIN}$$
$$n_e^{BIN}, \gamma_e^{BIN} \in \{0, 1\}$$

For example, if $E = \{0, 1, 2, 3\}$, the expression $2^0 \ n_0^{BIN} + 2^1 \ n_1^{BIN} + 2^2 \ n_2^{BIN} + 2^3 \ n_3^{BIN}$ can represent any integer number between 0 and 15. We can now replace the integer variables with their binary expansions:

$$n \ \mu = \mu \ \sum_{e \in E} 2^e \ n_e^{BIN}$$
$$\gamma \ \underline{\mu} = \underline{\mu} \ \sum_{e \in E} 2^e \ \gamma_e^{BIN}$$
$$\gamma \ \overline{\mu} = \overline{\mu} \ \sum_{e \in E} 2^e \ \gamma_e^{BIN}$$

Since the new bilinear terms are products of continuous and binary variables, the *McCormick Relaxation* can be applied: auxiliary continuous variables $z_e^n = \mu \ n_e^{BIN}$, $\underline{z}_e^\gamma = \underline{\mu} \ \gamma_e^{BIN}$, $\overline{z}_e^\gamma = \overline{\mu} \ \gamma_e^{BIN}$ are introduced so that

$$n \ \mu = \mu \ \sum_{e \in E} 2^e n_e^{BIN} = \sum_{e \in E} 2^e z_e^n$$
$$\gamma \ \underline{\mu} = \underline{\mu} \ \sum_{e \in E} 2^e \gamma_e^{BIN} = \sum_{e \in E} 2^e \underline{z}_e^\gamma$$
$$\gamma \ \overline{\mu} = \overline{\mu} \ \sum_{e \in E} 2^e \gamma_e^{BIN} = \sum_{e \in E} 2^e \ \overline{z}_e^\gamma$$

The following linear constraints link the variables

$$0 \leq \mu - z_e^n \leq M^\mu \ (1 - n_e^{BIN})$$
$$0 \leq z_e^n \leq M^\mu \ n_e^{BIN}$$
$$0 \leq \underline{\mu} - \underline{z}_e^\gamma \leq M^{\underline{\mu}} \ (1 - \gamma_e^{BIN})$$
$$0 \leq \underline{z}_e^\gamma \leq M^{\underline{\mu}} \ \gamma_e^{BIN}$$
$$0 \leq \overline{\mu} - \underline{z}_e^\gamma \leq M^{\overline{\mu}} \ (1 - \gamma_e^{BIN})$$
$$0 \leq \overline{z}_e^\gamma \leq M^{\overline{\mu}} \ \gamma_e^{BIN}$$

All the constants M are set to the $VoLL$. We defined different sets E for each integer variables to model the correct number of binary variables; for example, E_k^{K+KO} is the set of exponents of 2 required to represent the maximum number of ON thermal generators $(n_{k,0}^K+\overline{n}_k^K)$.

For the linearization of the product $Q\ \lambda$ we also applied the *piece-wise linearization* technique, later discarded because it had the disadvantages described below.

Following the technique applied in [Vespucci et al., 2013], we approximate the non-linear revenue terms $Q\ \lambda$ by piece-wise linear functions. For any real number Q and λ it holds that $Q\ \lambda = s^2 - d^2$, where

$$s = \frac{Q+\lambda}{2}\ ,\ d = \frac{Q-\lambda}{2}$$

The piecewise-linear approximations $\overline{s^2}$ of s^2 and $\overline{d^2}$ of d^2 are constructed as follows.

Q is defined in the interval $\left[0,\ \overline{Q}\right]$, where $\overline{Q}$ varies based on the technology considered (for example, for thermal plants $\overline{Q} = \overline{Q}_k^K\ (n_{k,0}^K + \overline{n}_k^K)$, according to Tab. 6.2 while λ is defined in $[0,\ VoLL]$, in general, or in $\left[0,\ \max_i \hat{\lambda}_i\right]$ more specifically; therefore, the bounds of s and d are:

$$s^{\min} = 0,\ \ s^{\max} = \frac{\overline{Q}+VoLL}{2};\quad d^{\min} = \frac{-VoLL}{2},\ \ d^{\max} = \frac{\overline{Q}}{2}$$

Let's $\hat{s}_b$, $b \in \mathcal{B}$, denote $|\mathcal{B}|$ breaking points chosen in $\left[s^{min}, s^{max}\right]$, with $\hat{s}_1 = s^{min}$ and $\hat{s}_{|\mathcal{B}|} = s^{max}$, so as to divide the interval into $|\mathcal{B}|-1$ subintervals. For example, with $|\mathcal{B}| = 5$, 4 subintervals of width $\Delta^s = (s^{max} - s^{min})/10$ are defined, and each of the 5 breakpoints is given by $\hat{s}_b = s^{min} + \Delta^s\ (b-1)$, with $b = 1, ..., 11$. Let's s_b^{coef}, $b \in \mathcal{B}$, denote variables belonging to a Special Order Set of type 2 (SOS2), i.e. a set of non-negative ordered variables of which at most two adjacent ones can take a positive value. The following linear constraints define the variable $\overline{s^2}$ as a piecewise-linear approximation of s^2:

$$s_b^{coef}\ \text{SOS2}, b \in \mathcal{B} \qquad \sum_{b\in\mathcal{B}} s_b^{coef} = 1$$

$$\frac{Q+\lambda}{2} = \sum_{b\in\mathcal{B}} \hat{s}_b\ s_b^{coef} \qquad \overline{s^2} = \sum_{b\in\mathcal{B}} \hat{s}_b^2\ s_b^{coef}$$

The piecewise-linear approximation $\overline{d^2}$ of d^2 is constructed analogously, yielding the constraints:

$$d_b^{coef} \text{ SOS2}, b \in \mathcal{B} \qquad \sum_{b \in \mathcal{B}} d_b^{coef} = 1$$

$$\frac{Q - \lambda}{2} = \sum_{b \in \mathcal{B}} \hat{d}_b \, d_b^{coef} \qquad \overline{d^2} = \sum_{b \in \mathcal{B}} \hat{d}_b^2 \, d_b^{coef}$$

Therefore, the revenues $Q \ \lambda = s^2 - d^2$ can be approximated by $\overline{s^2} - \overline{d^2}$. In Fig. 6.1, the values of s^2 (dashed parabola) and its approximation $\overline{s^2}$ (piecewise linear) are compared: $\overline{s^2}$ overestimates s^2. The overestimation error is $\epsilon s = \overline{s^2} - s^2$ and its magnitude vary depending on the position of the curve determined by the optimal values of the SOS2 variables s_b^{coef}.

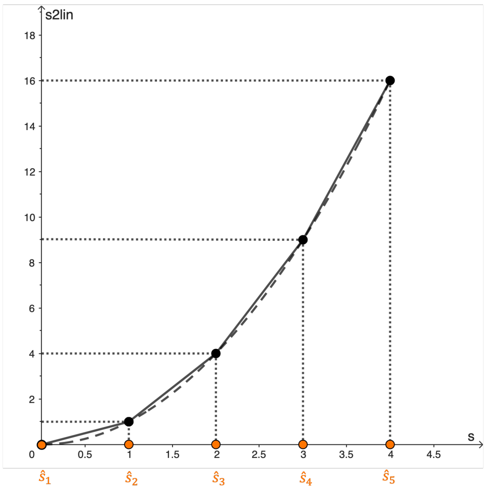


Figure 6.1: Comparison of s^2 (dashed parabola) and its piecewise linear approximation $\overline{s^2}$ (solid segments), considering the breaking points $\hat{s}_b, b \in \mathcal{B}$ (orange)

Since both $\overline{s^2}$ and $\overline{d^2}$ overestimate s^2 and d^2 respectively, the difference $\overline{s^2} - \overline{d^2}$ may turn out to be negative, even though the real (not approximated) revenue is always non-negative. This could lead to solutions being incorrectly considered infeasible when they are actually feasible.

Chapter 7

Computing feasible solutions and proving optimality

The computation of optimal solutions of GEP-RA models has proven to be a difficult task, due to the complexity arising from the two-level model structure, as well as the presence of bilinear terms and of binary and integer variables. In particular, two main difficulties have been observed, namely (i) obtaining a feasible solution of the GEP-RA models; (ii) proving optimality. This chapter describes some approaches we have introduced to overcome the above difficulties, which try to exploit the bilevel structure of the GEP-RA models.

7.1 Feasibility cuts, warm starts and heuristic callbacks

Commercial solvers, notably the best-performing Gurobi, struggle to find a feasible solution. The general-purpose heuristics used by Gurobi, such as the Relaxation Induced Neighborhood Search or the feasibility pump (see, e.g., [Fischetti and Lodi, 2010]), were not of much help in determining feasible solutions, especially when more than one representative day was considered in the GEP-RA model. Since the GEP-RA model is reformulated as a single-level optimization problem, general-purpose heuristics cannot detect its original special bilevel structure. We have therefore developed some heuristic approaches that take into account the special structure of the problem. Our efforts to improve performance have focused in particular on the GEP-RA (2) model, the UL and LL of which are reported below for the convenience of the reader.

UL

$$\min_{Var} \sum_{z\in Z}\sum_{t\in T} \lambda_{z,t}\ (D_{z,t} - B^{D}_{z,t} + B^{C}_{z,t}) + \sum_{z\in Z}(u^{V}_{z} + u^{WI}_{z} + u^{WO}_{z} + u^{B}_{z}) + \sum_{k\in K} u^{K}_{k} + \sum_{z\in Z}\sum_{t\in T} c^{RNP}\ RNP_{z,t} \quad (5.45)$$

Constraints on zonal capacities of RES power plants and batteries in the target year: $z \in Z$

$$\underline{V}_z \le \overline{Q}^{V}\ (n^{V}_{z,0} + n^{V}_{z}) \le \overline{V}_z \quad (5.4)$$

$$\underline{W}_z \le \overline{Q}^{WI}\ (n^{WI}_{z,0} + n^{WI}_{z}) + \overline{Q}^{WO}\ (n^{WO}_{z,0} + n^{WO}_{z}) \le \overline{W}_z \quad (5.5)$$

$$\underline{B}_z \le \overline{Q}^{B}\ (n^{B}_{z,0} + n^{B}_{z}) \le \overline{B}_z \quad (5.6)$$

Constraints on thermal power capacity in the target year: $k \in K_z,\ z \in Z$

$$n^{K}_{k} \le \overline{n}^{K}_{k} \quad (5.7)$$

UC constraints, with avoidance of physical withholding

$$\gamma_{k,t} - \gamma^{I}_{k,t} = \alpha_{k,t} - \beta_{k,t} \qquad k \in K, t \in T^{I} \quad (5.8)$$

$$\gamma_{k,t} - \gamma_{k,t-1} = \alpha_{k,t} - \beta_{k,t} \qquad k \in K, t \in T \setminus T^{I} \quad (5.9)$$

$$\gamma_{k,t} + \gamma^{SL}_{k,t} = n^{K}_{k,0} + n^{K}_{k} \qquad k \in K, t \in T \quad (5.10)$$

$$\gamma^{SL}_{k,t} \ge 1 - x^{TF}_{k,t} \qquad k \in K, t \in T \quad (5.11)$$

$$\gamma^{SL}_{k,t} \le (n^{K}_{k,0} + \overline{n}^{K}_{k})(1 - x^{TF}_{k,t}) \qquad k \in K, t \in T \quad (5.12)$$

$$x^{TF}_{k,t} \ge x^{TF}_{kk,t} \qquad q \in [1, |Q|-1]\,, k \in K_q, kk \in K_{q+1}, t \in T \quad (5.13)$$

Constraints on operation of batteries: $z \in Z$

$$B^{C}_{z,t} \le rt^{C}\ \overline{B}_z\ \phi_{z,t} \qquad t \in T \quad (5.15)$$

$$B^{D}_{z,t} \le D_{z,t}\ (1 - \phi_{z,t}) \qquad t \in T \quad (5.16)$$

$$B^{C}_{z,t} \le rt^{C}\ \overline{Q}^{B}\ (n^{B}_{z,0} + n^{B}_{z}) \qquad t \in T \quad (5.18)$$

$$B^{D}_{z,t} \le rt^{D}\ \overline{Q}^{B}\ (n^{B}_{z,0} + n^{B}_{z}) \qquad t \in T \quad (5.19)$$

$$B_{z,t} \le \overline{Q}^{B}\ (n^{B}_{z,0} + n^{B}_{z}) \qquad t \in T \quad (5.20)$$

$$B_{z,t} = (1-\kappa)\ B^{I}_{z,t} + \kappa^{C}\ B^{C}_{z,t} - \kappa^{D}\ B^{D}_{z,t} \qquad t \in T^{I} \quad (5.21)$$

$$B_{z,t} = (1-\kappa)\ B_{z,t-1} + \kappa^{C}\ B^{C}_{z,t} - \kappa^{D}\ B^{D}_{z,t} \qquad t \in T \setminus T^{I} \quad (5.22)$$

$$B_{z,t} = B^{I}_{z,t} \qquad t \in T^{L} \quad (5.23)$$

Reserve constraints: $z \in Z, t \in T$

$$\sum_{k\in K_z} (\overline{Q}^{K}_{k}\ \gamma_{k,t} - Q^{K}_{k,t}) + df \left[rt^{D}\ \overline{Q}^{B}\ (n^{B}_{z,0} + n^{B}_{z})\ (1 - \phi_{z,t}) - B^{D}_{z,t} \right] + RNP_{z,t} \ge R_{z,t} \quad (5.24)$$

Constraints that avoid price indeterminacy: $z \in Z,\ t \in T$

$$y_{z,t,1} \le \sum_{z'\in Z} Q^{V}_{z',t} \quad (5.25)$$

$$y_{z,t,2} \le \sum_{z'\in Z} Q^{WI}_{z',t} \quad (5.26)$$

$$y_{z,t,3} \le \sum_{z'\in Z} Q^{WO}_{z',t} \quad (5.27)$$

$$y_{z,t,3+q} \le \sum_{k\in K_q} \gamma_{k,t} \qquad q \in Q \quad (5.28)$$

$$\sum_{i\in I} y_{z,t,i} = 1 \quad (5.29)$$

$$\lambda_{z,t} = \sum_{i\in I} \hat{\lambda}_i\ y_{z,t,i} \quad (5.30)$$

Revenue adequacy constraints: $z \in Z$

$$\sum_{t\in T} \lambda_{z,t}\ Q^K_{k,t} + u^K_k \geq M \left[\frac{|T|}{8760}\ I^K_k\ \overline{Q}^K_k\ n^K_k + \sum_{t\in T} (C^M_k\ Q^K_{k,t} + C^{SU}_k\ \alpha_{k,t}) \right] \qquad k \in K_z \quad (5.31)$$

$$\sum_{t\in T} \lambda_{z,t}\ Q^V_{z,t} + u^V_z \geq M \frac{|T|}{8760}\ I^V_z\ \overline{Q}^V\ n^V_z \qquad (5.56)$$

$$\sum_{t\in T} \lambda_{z,t}\ Q^{WI}_{z,t} + u^{WI}_z \geq M \frac{|T|}{8760}\ I^{WI}_z\ \overline{Q}^{WI}\ n^{WI}_z \qquad (5.57)$$

$$\sum_{t\in T} \lambda_{z,t}\ Q^{WO}_{z,t} + u^{WO}_z \geq M \frac{|T|}{8760}\ I^{WO}_z\ \overline{Q}^{WO}\ n^{WO}_z \qquad (5.58)$$

$$\sum_{t\in T} \lambda_{z,t}\ (B^D_{z,t} - B^C_{z,t}) + u^B_z \geq M \frac{|T|}{8760}\ I^B_z\ \overline{Q}^B\ n^B_z \qquad (5.59)$$

LL (for each $t \in T$) (5.62)

$$\min_{Q,F_{l,t}} \sum_{z\in Z} \left(P^V\ Q^V_{z,t} + P^{WI}\ Q^{WI}_{z,t} + P^{WO}\ Q^{WO}_{z,t} \right) + \sum_{k\in K} P^K_k\ Q^K_{k,t} \qquad (5.42a)$$

$$Q^V_{z,t} + Q^{WI}_{z,t} + Q^{WO}_{z,t} + \sum_{k\in K_z} Q^K_{k,t} + B^D_{z,t} + \sum_{l\in BS_z} F_{l,t} \geq D_{z,t} + \sum_{l\in FS_z} F_{l,t} + B^C_{z,t} \qquad z \in Z \qquad (\lambda_{z,t} \geq 0) \quad (5.62b)$$

$$0 \leq Q^V_{z,t} \leq cf^V_{z,t}\ \overline{Q}^V\ (n^V_{z,0} + n^V_z) \qquad z \in Z \qquad (\mu^V_{z,t} \geq 0) \quad (5.42c)$$

$$0 \leq Q^{WI}_{z,t} \leq cf^{WI}_{z,t}\ \overline{Q}^{WI}\ (n^{WI}_{z,0} + n^{WI}_z) \qquad z \in Z \qquad (\mu^{WI}_{z,t} \geq 0) \quad (5.42d)$$

$$0 \leq Q^{WO}_{z,t} \leq cf^{WO}_{z,t}\ \overline{Q}^{WO}\ (n^{WO}_{z,0} + n^{WO}_z) \qquad z \in Z \qquad (\mu^{WO}_{z,t} \geq 0) \quad (5.42e)$$

$$\underline{Q}^K_k\ \gamma_{k,t} \leq Q^K_{k,t} \leq \overline{Q}^K_k\ \gamma_{k,t} \qquad k \in K \qquad (\underline{\mu}^K_{k,t} \geq 0, \overline{\mu}^K_{k,t} \geq 0) \quad (5.42f)$$

$$\underline{F}_l \leq F_{l,t} \leq \overline{F}_l \qquad l \in L \qquad (\underline{v}_{l,t} \geq 0, \overline{v}_{l,t} \geq 0) \quad (5.42g)$$

The first approach we propose consists of adding to the single-level reformulation the redundant constraint (7.2) for each hour t.

$$\sum_{z\in Z} \left[cf^V_{z,t}\ \overline{Q}^V\ (n^V_{z,0} + n^V_z) + cf^{WI}_{z,t}\ \overline{Q}^{WI}\ (n^{WI}_{z,0} + n^{WI}_z) + cf^{WO}_{z,t}\ \overline{Q}^{WO}\ (n^{WO}_{z,0} + n^{WO}_z) \right] +$$
$$+ \sum_{k\in K} \overline{Q}^K \gamma_{k,t} \geq \sum_{z\in Z} \left[D_{z,t} + B^C_{z,t} - B^D_{z,t} \right] \qquad (7.2)$$

This ensures that the system's production capacity, defined as a function of the UL variables ($n^V_z, n^{WI}_z, n^{WO}_z, \gamma_{k,t}, B^C_{z,t}, B^D_{z,t}$), which are treated as parameters in the LL, is sufficient to meet the total demand. In other words, it establishes a direct relation between the UL variables and the demand that the LL must meet, thus ensuring that the LL constraints are satisfied as long as the possible energy exchanges between zones do not conflict with the transmission capacity constraints (5.42g). Therefore, the constraints (7.2) can be seen as a kind of *feasibility cuts*.

The constraint (7.2) is derived as follows. The sum over $z \in Z$ of inequalities (5.62b) yields (7.3) since flows cancel out.

$$\sum_{z\in Z} \left(Q^V_{z,t} + Q^{WI}_{z,t} + Q^{WO}_{z,t} + \sum_{k\in K_z} Q^K_{k,t} \right) \geq \sum_{z\in Z} \left(D_{z,t} + B^C_{z,t} - B^D_{z,t} \right) \qquad (7.3)$$

Moreover, inequalities (5.42c)-(5.42f) imply the inequality

$$\sum_{z\in Z}\left[cf^V_{z,t}\,\overline{Q}^V\,(n^V_{z,0}+n^V_z)+cf^{WI}_{z,t}\,\overline{Q}^{WI}\,(n^{WI}_{z,0}+n^{WI}_z)+cf^{WO}_{z,t}\,\overline{Q}^{WO}\,(n^{WO}_{z,0}+n^{WO}_z)\right]+$$
$$+\sum_{k\in K}\overline{Q}^K\gamma_{k,t}\geq\sum_{z\in Z}\left(Q^V_{z,t}+Q^{WI}_{z,t}+Q^{WO}_{z,t}+\sum_{k\in K_z}Q^K_{k,t}\right) \tag{7.4}$$

and from (7.3) and (7.4) inequality (7.2) follows.

We note that when using Gurobi, (7.2) for $t\in T$ can be introduced as lazy constraints with option 3 that cuts off the relaxation solution at the root node if the constraint is violated.

The second approach develops heuristic algorithms that serve a twofold purpose: they are designed not only to compute an initial feasible solution, i.e. to provide a warm start, but also to attempt to enhance the solution at any node within the branch-and-cut process, i.e. to act in a callback setting.

To provide a concise description of the heuristic procedures, we define the following partitions of the variable set from the GEP-RA (2), which represent the warm start values.

$$Val_A=\{n^B_z,n^K_k,\phi_{z,t},\alpha_{k,t},\beta_{k,t},\gamma^{SL}_{k,t},x^{TF}_{k,t},B_{z,t},RNP_{z,t}\}$$
$$Val_B=\{n^V_z,n^{WI}_z,n^{WO}_z,\gamma_{k,t},B^D_{z,t},B^C_{z,t}\}$$
$$Val_C=\{Q^V_{z,t},Q^{WI}_{z,t},Q^{WO}_{z,t},Q^K_{k,t},F_{l,t}\}$$
$$Val_D=\{\lambda_{z,t},\mu^V_{z,t},\mu^{WI}_{z,t},\mu^{WO}_{z,t},\underline{\mu}^K_{k,t},\overline{\mu}^K_{k,t},\underline{\upsilon}_{l,t},\overline{\upsilon}_{l,t}\}$$
$$Val_E=\{u^V_z,u^{WI}_z,u^{WO}_z,u^B_z,u^K_k\}$$
$$Val_F=\{y_{z,t,i}\}$$

In the developed software code, we assign a binary parameter to each variable within the problem (1 for the variable to be warm started, 0 otherwise), so that the user can activate a warm start for any combination of selected variables, initializing them according to their respective values Val. This relies on Gurobi's ability to handle partially populated warm starts by optimizing with respect to the variables that are not warm started. The success of Gurobi in completing the solution depends on which variables are initialized and on the number of branch-and-cut nodes explored when completing a partial start, which can be set with the parameter `StartNodeLimit`.

The heuristic procedures determine feasible solutions by mimicking the hierarchical relationship between the CP, in the UL, and the MO, in the LL. The bilevel configuration described in Fig 4.8 is recalled: the UL determines the values for RES investments, thermal statuses, and battery charging and discharging (referred to as Val_B); these values are received by the LL, which returns the optimal values of accepted quantities (referred to as Val_C) and prices (referred to as Val_D). However, assigning values to Val_B through a naive approach, as in the first attempt of Algorithm (6), can lead to an infeasible solution for the LL. In particular, Algorithm (6) sets the number of new photovoltaic generators to its maximum, as defined by the constraint (5.4). The number of new offshore wind generators is set to zero and the number of inshore wind generators is set to its maximum, as defined by the constraint (5.5), which limits the sum of inshore and offshore wind capacity. Therefore, the maximum value of RES production can be calculated accordingly. For each zone z, the set K_z containing the thermal generators within the zone, is sorted in ascending order of offer price P_k^K. This prioritizes generators with lower offer prices for operational decisions, thus avoiding physical withholding as required by the original model. For each hour t, the status $\gamma_{k,t}$ of the generator k is initialized to zero and then updated until the capacity is sufficient to cover the zonal hourly demand (net of RES production). Setting both the battery charging and discharging to zero may lead to an infeasible solution when dealing with the LL. This happens in test cases where the batteries are needed to cover the demand. To overcome this issue, we could set the battery discharges to match the capacity deficit, calculated as $B_{z,t}^D = D_{z,t} - prod_z^{RES} - \sum_{k \in K_z} \overline{Q}^K \gamma_{k,t}$. However, in turn, it is difficult to determine the related charges $B_{z,t}^C$ and at the same time, take into account

Algorithm 6: Naive Warm start

Data: An instance of the GEP-RA

Result: Val_B, Val_C, Val_D // Val_A, Val_E, Val_F set by Gurobi (partial WS)

begin

- Set $n_z^V = \frac{\overline{V}_z}{\overline{Q}^V} - n_{z,0}^V$ // max solar
- Set $n_z^{WO} = 0$; $n_z^{WI} = \frac{\overline{W}_z}{\overline{Q}^{WI}} - n_{z,0}^{WI}$ // max wind
- $prod_z^{RES} = cf_{z,t}^V \, \overline{Q}^V \, (n_{z,0}^V + n_z^V) + cf_{z,t}^{WI} \, \overline{Q}^{WI} \, (n_{z,0}^{WI} + n_z^{WI})$
- **for** $z \in Z$ **do**
  - Order K_z // according to P_k^K
  - **for** $t \in T$ **do**
    - **for** $k \in K_z$ **do**
      - Set $\gamma_{k,t} = 0$
      - **while** $\sum_{kk \in K_z} \overline{Q}^K \gamma_{kk,t} < D_{z,t} - prod_z^{RES}$ **do**
        - **if** $\gamma_{k,t} < n_{k,0}^K + \overline{n}_k^K$ **then**
          - $\gamma_{k,t} \leftarrow \gamma_{k,t} + 1$
- Set $B_{z,t}^D = B_{z,t}^C = 0$
- $Val_C \leftarrow$ Solve LL (5.62) // considering Val_B as defined above
- $Val_D \leftarrow$ Marginals of LL constraints

flows that can be used to meet demand in one zone with the excess capacity of another zone, according to the capacity of the lines and to the price signals.

The difficulties outlined underscore the necessity of a more principled approach. Instead of a procedure that assigns the values Val_B manually with, e.g., the logic above, this more formal approach can be implemented via the formulation of a specific auxiliary optimization problem. We therefore design this auxiliary problem such that it determines values Val_B that are both feasible for the LL and aligned with UL constraints in terms of battery dynamics.

The idea of the proposed mixed-integer model AUX 1 (7.5) is to hybridize variables and constraints from both the UL and the LL of GEP-RA (2). Starting from the UL formulation, AUX 1 therefore considers constraints related to investment capacity (7.5b), operation of both thermal plants (7.5c) and batteries (7.5d), and reserve requirement (7.5e). Given the final goal of warm starting the GEP-RA (2), AUX 1 does not take into account market clearing prices, so we do not consider the RA constraints (5.31), (5.56)-(5.59) and the constraints designed to prevent price indeterminacy (5.25)-(5.30). This is possible because the presence of side payments in the original GEP-RA (2) model ensures that the RA conditions can be met, irrespective of the revenues determined by the LL. Furthermore, the price determination is allocated to the LL, which operates independently of AUX 1. This separation effectively eliminates the potential for decisions within AUX 1 to strategically induce price indeterminacy. AUX 1 is also designed to include constraints (7.5f)-(7.5k) to ensure that the optimal values Val_B are feasible for the LL. In particular, (7.5f) ensures that investments in RES and the operation of thermal plants and batteries determine a capacity sufficient to meet demand, (7.5g)-(7.5j) ensures consistency between investments and quantities, while the flows contributing to the demand constraint are limited by the capacities of the lines through (7.5k). Note that these constraints are identical to the LL constraints, but $n_z^V, n_z^{WI}, n_z^{WO}, \gamma_{k,t}, B_{z,t}^D, B_{z,t}^C$ are variables instead of parameters. Moreover, the optimal values for the quantities determined by AUX 1 are not considered by the complete warm start algorithm, to be presented shortly, since they may differ from those determined by LL, which is a linear problem since $n_z^V, n_z^{WI}, n_z^{WO}, \gamma_{k,t}, B_{z,t}^D, B_{z,t}^C$ are fixed to Val_B. Starting from (5.45), the objective function is formulated as (7.5a). This revised formulation retains the focus on minimizing unmet reserve requirements and costs, now expressed in terms of quantities rather than prices, but excludes minimizing side payments. Lastly, we observe that the problem (7.5) can be solved in a few seconds.

$$\min_{Var_1} \sum_{t\in T}\left[\sum_{z\in Z}\left(P^V\ Q^V_{z,t} + P^{WI}\ Q^{WI}_{z,t} + P^{WO}\ Q^{WO}_{z,t}\right) + \sum_{k\in K} P^K_k\ Q^K_{k,t}\right] + \\ + \sum_{t\in T}\sum_{z\in Z} c^{RNP}\ RNP_{z,t} \tag{7.5a}$$

$$(5.4) - (5.7) \tag{7.5b}$$

$$(5.8) - (5.13) \tag{7.5c}$$

$$(5.15), (5.16), (5.18) - (5.23) \tag{7.5d}$$

$$(5.24) \tag{7.5e}$$

$$Q^V_{z,t} + Q^{WI}_{z,t} + Q^{WO}_{z,t} + \sum_{k\in K_z} Q^K_{k,t} + B^D_{z,t} + \sum_{l\in BS_z} F_{l,t} \geq D_{z,t} + \sum_{l\in FS_z} F_{l,t} + B^C_{z,t} \qquad z\in Z \tag{7.5f}$$

$$0 \leq Q^V_{z,t} \leq cf^V_{z,t}\ \overline{Q}^V\ (n^V_{z,0} + n^V_z) \qquad z\in Z \tag{7.5g}$$

$$0 \leq Q^{WI}_{z,t} \leq cf^{WI}_{z,t}\ \overline{Q}^{WI}\ (n^{WI}_{z,0} + n^{WI}_z) \qquad z\in Z \tag{7.5h}$$

$$0 \leq Q^{WO}_{z,t} \leq cf^{WO}_{z,t}\ \overline{Q}^{WO}\ (n^{WO}_{z,0} + n^{WO}_z) \qquad z\in Z \tag{7.5i}$$

$$\underline{Q}^K_k\ \gamma_{k,t} \leq Q^K_{k,t} \leq \overline{Q}^K_k\ \gamma_{k,t} \qquad k\in K \tag{7.5j}$$

$$\underline{F}_l \leq F_{l,t} \leq \overline{F}_l \qquad l\in L \tag{7.5k}$$

$Var_1 \equiv$

$$\begin{aligned}
& n^V_z, n^{WI}_z, n^{WO}_z, n^B_z \in \mathbb{N} && z\in Z \\
& n^K_k \in \mathbb{N} && k\in K \\
& \alpha_{k,t}, \beta_{k,t}, \gamma_{k,t}, \gamma^{SL}_{k,t} \in \mathbb{N} && k\in K, t\in T \\
& x^{TF}_{k,t} \in \{0,1\} && k\in K, t\in T \\
& B_{z,t}, B^C_{z,t}, B^D_{z,t}, RNP_{z,t} \geq 0 && z\in Z, t\in T \\
& \phi_{z,t} \in \{0,1\} && z\in Z, t\in T \\
& Q^V_{z,t}, Q^{WI}_{z,t}, Q^{WO}_{z,t} \geq 0 && z\in Z, t\in T \\
& Q^K_{k,t} \geq 0 && k\in K, t\in T \\
& F_{l,t} \text{ free} && l\in L, t\in T
\end{aligned} \tag{7.5l}$$

The complete Algorithm (7) considers the auxiliary problem AUX 1 for computing values Val_A and Val_B; then Val_B are considered as parameters of the LL, which determines Val_C

and Val_D. The values Val_E and Val_F are computed ex-post. Specifically, the side payments are calculated as the difference between market revenues, which are a function of quantities and prices (determined by LL), and investment and operational costs, which are a function of investment and operational decisions (determined by AUX 1). The binary parameters $y_{z,t,i}$ are determined to identify which among the offer prices $\hat{\lambda}_i$ has been selected as the market clearing price $\lambda_{z,t}$ by the LL.

Algorithm 7: Warm start and Heuristic callback

Data: An instance of the GEP-RA
Result: Val

begin
  $Val_A, Val_B \leftarrow$ `Solve` AUX 1 (7.5)
  $Val_C \leftarrow$ `Solve` LL (5.62) `// considering` Val_B `from AUX 1`
  $Val_D \leftarrow$ Marginals of LL constraints

$$Val_E \leftarrow u_z^V = \sum_{t\in T} \lambda_{z,t}\ Q_{z,t}^V - M\ \frac{|T|}{8760}\ I_z^V\ \overline{Q}^V\ n_z^V$$
$$u_z^{WI} = \sum_{t\in T} \lambda_{z,t}\ Q_{z,t}^{WI} - M\ \frac{|T|}{8760}\ I_z^{WI}\ \overline{Q}^{WI}\ n_z^{WI}$$
$$u_z^{WO} = \sum_{t\in T} \lambda_{z,t}\ Q_{z,t}^{WO} - M\ \frac{|T|}{8760}\ I_z^{WO}\ \overline{Q}^{WO}\ n_z^{WO}$$
$$u_z^B = \sum_{t\in T} \lambda_{z,t}\ (B_{z,t}^D - B_{z,t}^C) - M \frac{|T|}{8760}\ I_z^B\ \overline{Q}^B\ n_z^B$$
$$u_k^K = \sum_{t\in T} \lambda_{z,t}\ Q_{k,t} - M \left[\frac{|T|}{8760}\ I_k^K\ \overline{Q}_k^K\ n_k^K + \sum_{t\in T}(C_k^M\ Q_{k,t} + C_k^{SU}\ \alpha_{k,t}) \right]$$
$$Val_F \leftarrow y_{z,t,i} : \lambda_{z,t} = \sum_{i\in I} \hat{\lambda}_i\ y_{z,t,i}$$

Note that a complete warm start of GEP-RA (2), i.e. a warm start of all variables, including the values $RNP_{z,t}$ for unmet reserve from AUX 1 and the values $Q_{k,t}^K$ for thermal accepted quantities from LL, may lead to an infeasible solution and therefore cause the warm start to be rejected by Gurobi. In fact, as mentioned above, the quantities determined by AUX 1 can differ from those determined by the LL and thus violate (5.24). An example of the variables that can be initialized to obtain a feasible warm start is given in Section (8.7).

In the enhanced proposal presented in Algorithm (8), an alternative approach within the same framework renders AUX 1 more adherent to the UL. Starting from AUX 1 (7.5), AUX 2 (7.6) integrates RA constraints (7.6c)-(7.6g) with *fixed prices from the LL*. Moreover, the objective function, now modified as in (7.6a), now incorporates side payments and the product of fixed prices for charging and discharging the batteries. In this way, (7.6) leads to more consistent values for the same variables since side payments are included in the optimization and not computed ex-post. The algorithm begins by setting all zonal clearing prices to the minimum offer price, which is feasible for AUX 2 due to the inclusion of side payments in the RA constraints. Prices are then differentiated based on zone and time through the resolution of LL, and subsequently employed to solve AUX 2 once more. It is noteworthy that the LL returns prices to AUX 2 solely because the quantities would restrict both investments and operations.

Algorithm 8: Warm start and Heuristic callback - alternative version

Data: An instance of the GEP-RA
Result: Val

begin
- `Set` $\lambda_{z,t} = \min_{i \in I} \hat{\lambda}_i$
- $Val_B \leftarrow$ `Solve` AUX 2 (7.6)
- $Val_C \leftarrow$ `Solve` LL (5.62) — `// considering` Val_B `from AUX 2`
- $Val_D \leftarrow$ Marginals of LL constraints
- $Val_A, Val_B, Val_E \leftarrow$ `Solve` AUX 2 (7.6) — `// considering` $\lambda_{z,t}$ `from LL`
- $Val_F \leftarrow y_{z,t,i} : \lambda_{z,t} = \sum_{i \in I} \hat{\lambda}_i \; y_{z,t,i}$

$$\min_{Var_2} \sum_{t \in T} \left[\sum_{z \in Z} \left(P^V \; Q^V_{z,t} + P^{WI} \; Q^{WI}_{z,t} + P^{WO} \; Q^{WO}_{z,t} \right) + \sum_{k \in K} P^K_k \; Q^K_{k,t} \right] +$$
$$+ \sum_{t \in T} \sum_{z \in Z} c^{RNP} \; RNP_{z,t} +$$
$$+ \sum_{z \in Z} (u^V_z + u^{WI}_z + u^{WO}_z + u^B_z) + \sum_{k \in K} u^K_k + \sum_{z \in Z} \sum_{t \in T} \lambda_{z,t} \; (B^C_{z,t} - B^D_{z,t}) \tag{7.6a}$$

$$(7.5b) - (7.5k) \tag{7.6b}$$

$$\sum_{t \in T} \lambda_{z,t} \; Q^K_{k,t} + u^K_k \geq M \left[\frac{|T|}{8760} \; I^K_k \; \overline{Q}^K_k \; n^K_k + \sum_{t \in T} (C^M_k \; Q^K_{k,t} + C^{SU}_k \; \alpha_{k,t}) \right] \quad k \in K_z, z \in Z \tag{7.6c}$$

$$\sum_{t \in T} \lambda_{z,t} \; Q^V_{z,t} + u^V_z \geq M \frac{|T|}{8760} \; I^V_z \; \overline{Q}^V \; n^V_z \quad z \in Z \tag{7.6d}$$

$$\sum_{t \in T} \lambda_{z,t} \; Q^{WI}_{z,t} + u^{WI}_z \geq M \frac{|T|}{8760} \; I^{WI}_z \; \overline{Q}^{WI} \; n^{WI}_z \quad z \in Z \tag{7.6e}$$

$$\sum_{t \in T} \lambda_{z,t} \; Q^{WO}_{z,t} + u^{WO}_z \geq M \frac{|T|}{8760} \; I^{WO}_z \; \overline{Q}^{WO} \; n^{WO}_z \quad z \in Z \tag{7.6f}$$

$$\sum_{t \in T} \lambda_{z,t} \; (B^D_{z,t} - B^C_{z,t}) + u^B_z \geq M \frac{|T|}{8760} \; I^B_z \; \overline{Q}^B \; n^B_z \quad z \in Z \tag{7.6g}$$

$$
\begin{aligned}
Var_2 \equiv & Var_1 (7.5l) \cup \\
& u_z^V, u_z^{WI}, u_z^{WO}, u_z^B \geq 0 && z \in Z \\
& u_k^K \geq 0 && k \in K \quad (7.6h)
\end{aligned}
$$

It should be noted that a complete warm start downstream of Algorithm (8) may not be feasible. In fact, the accepted quantities determined by AUX 2 (to satisfy the RA constraints) could differ from those determined by the LL (and related to the fixed prices). As the side payments are determined by AUX 2, a warm start of both the quantities from the LL and the side payments from AUX 2 could result in an infeasible warm start, with the RA constraints potentially being violated.

Algorithm (8) can be developed into an iterative approach where Val_B are returned to LL for computing new $\lambda_{z,t}$ and so on. It should be noted that there is no guarantee of a decrease in the objective function of AUX 2, nor can there be any proof of convergence to optimal values of the whole problem.

Both algorithms (7) and (8) can also be invoked via a heuristic callback after some iterations, thus inside the branch-and-cut process. This procedure shares the same underlying logic but may simplify the task of completing the partial solution and lead to a solution that is closer to the optimum. In addition, the use of callbacks makes it possible to experiment with fixing different combinations of variables during the optimization process while the warm start can be only performed with one single combination at a time.

Additionally, we examine the alternative heuristic callback approach proposed in [Goyal and Richard, 2024], which considers the relaxed solution at each node of the branch-and-cut tree of a bilevel problem with binaries in the UL. If the values of the UL binaries in the relaxed solution exceed a certain threshold, the heuristic rounds them to 1. Conversely, if the values fall below the threshold, the heuristic sets them to 0. This rounding is then complemented by adjusting the associated continuous variables for consistency. Once a feasible UL solution has been obtained, the heuristic proceeds to solve both the primal and the dual of the LL. Although this method benefits from updating the callback values at different nodes, it falls short in our context, where securing feasible solutions for both the UL, which contains integer variables, and the LL presents significant obstacles. In our case, the heuristic rounding procedure results in a number of generators that either exceeds the limits set by the investment constraints (when rounding up) or is insufficient to meet demand (when rounding down). Moreover, rounding could lead to a solution that does not satisfy the RA constraints.

7.2 Valid inequalities for enhancing the CSC

One of the main reasons why the solver struggles to understand the interplay between levels is that, in the single-level reformulation determined with the Big-M or the SOS1 approaches, the dual variables of the LL are loosely coupled or even fully decoupled from the primal upper and lower level variables. In particular, at each node of the branch-and-cut process constructed by the solver, the binary variables of type $b^{CC} \in \{0,1\}$ of the big-M constraints (A.5a)-(A.5q) are relaxed to be continuous $b^{CC} \in [0,1]$, thus relaxing the optimality of the LL. Using the SOS1 reformulation, the conditions (A.4a)-(A.4q) are completely omitted at the beginning of the solution process, thus ignoring the optimality of the LL, and only later reintroduced via branching. This results in weak (initial) relaxations, indicated by low values for the best bound in the context of minimization, both at the root node and at each subsequent node in the branch-and-cut process. It should be noted that the term 'branch-and-cut' is used because the solvers automatically introduce their own cuts after the root node, thus not performing a pure branch-and-bound. However, these cuts are not bilevel-specific and thus may fall short in proving optimality. The authors of [Kleinert, 2021] present general cuts that are derived from strong duality conditions to *partially* address the issue of missing coupling. The reasons behind this partial coupling will be clarified shortly. Following their approach, for each $t \in T$ the Strong Duality condition (6.3) is written in the equivalent form (7.7). This form is equivalent because for the weak duality (which holds when the primal and dual solutions are both feasible) is $LLD \leq LLP$ and therefore the inequality (7.7) $LLD \geq LLP$ can only be satisfied at equality.

$$
\begin{aligned}
&\sum_{z\in Z} D_{z,t}\lambda_{z,t} + \sum_{l\in L}\left(\underline{F}_l\ \underline{\upsilon}_{l,t} - \overline{F}_l\ \overline{\upsilon}_{l,t}\right) \\
&- \sum_{z\in Z}\left[cf^{V}_{z,t}\ \overline{Q}^{V}\ n^{V}_{z,0}\ \mu^{V}_{z,t} + cf^{WI}_{z,t}\ \overline{Q}^{WI}\ n^{WI}_{z,0}\ \mu^{WI}_{z,t} + cf^{WO}_{z,t}\ \overline{Q}^{WO}\ n^{WO}_{z,0}\ \mu^{WO}_{z,t}\right] \\
&+ \sum_{z\in Z}(-B^{D}_{z,t}\ \lambda_{z,t} + B^{C}_{z,t}\ \lambda_{z,t}) \\
&- \sum_{z\in Z}\left[cf^{V}_{z,t}\ \overline{Q}^{V}\ n^{V}_{z}\ \mu^{V}_{z,t} + cf^{WI}_{z,t}\ \overline{Q}^{WI}\ n^{WI}_{z}\ \mu^{WI}_{z,t} + cf^{WO}_{z,t}\ \overline{Q}^{WO}\ n^{WO}_{z}\ \mu^{WO}_{z,t}\right] \\
&+ \sum_{k\in K}\left(\underline{Q}^{K}_{k}\ \gamma_{k,t}\ \underline{\mu}^{K}_{k,t} - \overline{Q}^{K}_{k}\ \gamma_{k,t}\ \overline{\mu}^{K}_{k,t}\right) \geq \\
&\qquad\qquad \sum_{z\in Z}\left(P^{V}Q^{V}_{z,t} + P^{WI}Q^{WI}_{z,t} + P^{WO}Q^{WO}_{z,t}\right) + \sum_{k\in K} P^{K}_{k}Q^{K}_{k,t} \qquad (7.7)
\end{aligned}
$$

The SD (7.7) is then used to derive the *valid inequality* (7.8) that can be added if violated either in the root node or in the branch-and-cut tree when using the SOS1 or the Big-M as the main reformulation of the CSC. For the derivation of (7.8), the bilinear products of (7.7) are relaxed by replacing each UL variable with valid approximations: the resulting left-hand side expression of (7.8) overestimates the left-hand side of (7.7) so that the linearized constraint remains valid. For example, $B^D_{z,t}\ \lambda_{z,t}$ is replaced by $B^{D\ -}_{z,t}\ \lambda_{z,t}$, where $B^{D\ -}_{z,t}$ is the minimum value the discharge can take in zone z, in hour t. Conversely, $B^C_{z,t}\ \lambda_{z,t}$ is replaced by $B^{C\ +}_{z,t}\ \lambda_{z,t}$, where $B^{C\ +}_{z,t}$ is the maximum value the charge can take in zone z, in hour t. Also the approximation of the number of RES generators, e.g. $n^{V\ -}_{z,t}$, depends on the time t since it is related to the hourly capacity factor. The approximated values are computed with the auxiliary MIQCP problems (7.9)-(7.14), which are derived from each of the primal constraints of the LL. The shared constraint set Ω is the set of both UL (including the RA constraints with bilinear terms) and LL-primal constraints, while Ω_D is the set of LL-dual constraints. The CSC are ignored in the auxiliary problems.

$$
\begin{aligned}
&\sum_{z\in Z} D_{z,t}\lambda_{z,t} + \sum_{l\in L}\left(\underline{F}_l\ \underline{\upsilon}_{l,t} - \overline{F}_l\ \overline{\upsilon}_{l,t}\right)\\
&-\sum_{z\in Z}\left[cf^V_{z,t}\ \overline{Q}^V\ n^V_{z,0}\ \mu^V_{z,t} + cf^{WI}_{z,t}\ \overline{Q}^{WI}\ n^{WI}_{z,0}\ \mu^{WI}_{z,t} + cf^{WO}_{z,t}\ \overline{Q}^{WO}\ n^{WO}_{z,0}\ \mu^{WO}_{z,t}\right]\\
&+\sum_{z\in Z}(-B^{D\ -}_{z,t}\ \lambda_{z,t} + B^{C\ +}_{z,t}\ \lambda_{z,t})\\
&-\sum_{z\in Z}\left[cf^V_{z,t}\ \overline{Q}^V\ n^{V\ -}_{z,t}\ \mu^V_{z,t} + cf^{WI}_{z,t}\ \overline{Q}^{WI}\ n^{WI-}_{z,t}\ \mu^{WI}_{z,t} + cf^{WO}_{z,t}\ \overline{Q}^{WO}\ n^{WO-}_{z,t}\ \mu^{WO}_{z,t}\right]\\
&+\sum_{k\in K}\left(\underline{Q}^K_k\ \gamma^+_{k,t}\ \underline{\mu}^K_{k,t} - \overline{Q}^K_k\ \gamma^-_{k,t}\ \overline{\mu}^K_{k,t}\right) \geq\\
&\qquad\qquad \sum_{z\in Z}\left(P^V Q^V_{z,t} + P^{WI}Q^{WI}_{z,t} + P^{WO}Q^{WO}_{z,t}\right) + \sum_{k\in K} P^K_k Q^K_{k,t}
\end{aligned}
\tag{7.8}
$$

It is noteworthy that the cut (7.8) couples the dual variables of the LL to the primal LL variables, while still leaving the UL variables decoupled.

for each $z \in Z, t \in T$

$$B^{D\ -}_{z,t}, B^{C\ +}_{z,t} := \min_{Var} B^D_{z,t} - B^C_{z,t} \tag{7.9a}$$

$$Var \in \Omega \times \Omega_D \tag{7.9b}$$

$$n^{V\ -}_{z,t} := \min_{Var} n^V_z \tag{7.10a}$$

$$Var \in \Omega \times \Omega_D \tag{7.10b}$$

$$n_{z,t}^{WI^-} := \min_{Var} n_z^{WI} \tag{7.11a}$$

$$Var \in \Omega \times \Omega_D \tag{7.11b}$$

$$n_{z,t}^{WO^-} := \min_{Var} n_z^{WO} \tag{7.12a}$$

$$Var \in \Omega \times \Omega_D \tag{7.12b}$$

for each $k \in K, t \in T$

$$\gamma_{k,t}^+ := \max_{Var} \gamma_{k,t} \tag{7.13a}$$

$$Var \in \Omega \times \Omega_D \tag{7.13b}$$

$$\gamma_{k,t}^- := \min_{Var} \gamma_{k,t} \tag{7.14a}$$

$$Var \in \Omega \times \Omega_D \tag{7.14b}$$

In our applications, most of the approximate values that enter the cut (7.8) with a negative sign are zero and therefore result in ineffective cuts. In fact, evaluating the left-hand side (for the LL-dual variables, since the UL variables are approximated) and the right-hand side with the current values through a callback at any node results in the left-hand side being already greater than the right-hand side. Therefore, when checking for cut violations, this is never violated. The reasons behind this are given by the following considerations. First of all, in the optimization problem (7.9), each hour is treated independently, thereby neglecting the dynamics of charge and discharge between consecutive hours. The minimization of the discharge for each hour t leads to zero discharge values ($B_{z,t}^{D^-} = 0$), except in hours and zones where discharge is necessary to meet the demand and avoid infeasibility. The maximization of the charge leads to the same values for all the hours ($rt^C\overline{B}_z$ for (5.6) and (5.18)), except for the last hour of the representative day, where the level of the battery must be equal to the initial level for (5.23). Moreover, in the optimization problems (7.11) and (7.12), the constraint (5.5) requires that the combined wind capacity from both inshore and offshore sources is greater than a minimum threshold. When minimizing inshore, the constraint (5.5) is satisfied by offshore and vice versa for the minimization of offshore wind generators. This interplay leads to zero minimum values for both wind type ($n_{z,t}^{WI-} = 0, n_{z,t}^{WO-} = 0$). Furthermore, in the optimization problem (7.14), the minimization of the number of on-line thermal generators leads to zero values ($\gamma_{k,t}^{-} = 0$), except in hours when the thermal capacity is necessary to satisfy the demand or the reserve.

We note that the good results of [Kleinert et al., 2021b] are obtained with a large data set of bilevel problems with both linear UL and LL, while our UL is strongly non-convex because it contains bilinear terms and integer variables. In addition, the authors use the CPLEX version 12.10 as the solver as well as a customized branch-and-bound and it is possible that Gurobi 11, which we use, is already much more efficient.

7.3 Hybrid CSC

Starting from the original formulation of the CSC, each reformulation introduced in Section (6.1) represents a (quite) different method of imposing the optimality of the LL. Recalling the work of [Kleinert, 2021] described in the previous Section (7.2), which uses the SOS1 reformulation and introduces a linear cut derived from the SD, the proposal that we present here is more general and leverages the Gurobi capabilities of handling bilinear components.

In this spirit, we envisioned and implemented a *hybrid approach* that combines pairs, or triplets, of selected formulations for the CSC. For example, we consider the Big-M reformulation (A.5a)-(A.5q) in conjunction with the original Product formulation (6.1a)-(6.1l), treating the constraints of the former as redundant. It is important to note that in the context of the present discussion, the term “redundant” takes on a specific meaning. In general, the integration of redundant constraints has been studied in the literature, see e.g. [Ruiz and Grossmann, 2011] and [Lalla-Ruiz and Voss, 2016] for examples in a mixed integer linear setting. However, to the best of our knowledge, this “hybridization” technique has never been proposed in the context of CSC reformulation within a bilevel optimization setting. Redundant constraints can introduce an extra level of pruning, propagation activity, cut generation, etc., which in turn can tighten the LP relaxation of the original problem, improve the quality of the lower bound (in the case of minimization), and produce better primal solutions, thus speeding up the overall execution.

If we consider the Big-M reformulation logic itself, we note that it is limited to branching on a single Big-M constraint and does not effectively connect UL variables with LL variables (both primal and dual) in solving the relaxed problem. In fact, in the LP relaxation, the binaries can take fractional values and consequently do not strictly impose that the product is null. We conjecture that this can be done instead by, e.g., the Product constraints in specific moments of the branch-and-cut procedure. When the solver branches on the binaries of the Big-M, i.e. fixes the binary variables in the branch-and-cut, it simultaneously fixes to zero one of the two terms of the Product constraints. Vice versa the (internal) reformulation of the Product constraints can lead to a fixing of some binary variables of the Big M formulation.

A third CSC reformulation, given by the SD (A.6), offers an additional, quite different, perspective on the relations between variables. Moreover, in the SD reformulation, the bilinear components differ from those of the product mode. We surmise that this ensemble can further enrich the whole formulation in the spirit discussed above. The computational tests presented later in (8.7) somehow confirm this belief.

In [Kleinert, 2021], a *linear* user cut is derived from the SD in order to allow CPLEX to solve the problem, as CPLEX is not able to deal with bilinearities. This linearization, however, results in the loss of the relationships between lower and upper-level variables. In contrast, our approach exploits Gurobi's ability to handle the bilinear terms of the Product formulation and/or of the SD, maintaining these variable relationships. It might be tempting to attribute our improved results solely to Gurobi's ability to manage bilinearities of the Product formulation or to the addition of upper bound constraints to each CSC product term as follows, mirroring the bounds given by the coefficients of the Big-M reformulation.

$$cp \leq M^{cp} \qquad vd \leq M^{vd} \tag{7.15a}$$

$$cd \leq M^{cd} \qquad vp \leq M^{vp} \tag{7.15b}$$

However, as will be demonstrated in the computational results presented in Section (8.7), merely relying on formulating CSC as products or on adding these bounds is insufficient for obtaining good dynamics in terms of the optimality gap in a reasonable time. This underscores the importance of combining CSC formulations to achieve better computational efficiency.

In order to further enhance the effectiveness of the proposed hybrid approach, it would be beneficial to investigate which constraints of the combination of the CSC reformulation should be written in redundant form. In the proposed approach, we have indeed omitted this step, leaving all constraints. It would also be of interest to examine the most appropriate reformulations/linearizations for each of these constraints.

Linearization approaches could be employed to handle bilinear terms and ascertain the efficacy of the hybrid formulation without relying on Gurobi. The work [Bestuzheva et al., 2023] describes a novel Reformulation Linearization Technique (RLT) cut separation, which appears to have been implemented in Gurobi. RLT is a deeply analyzed methodology for constructing tight linear relaxations of non-convex continuous and mixed-integer optimization problems. It was initially proposed in the seminal papers [Sherali and Adams, 1990], [Sherali and Adams, 1999]. Specifically, RLT constructs relaxations by first multiplying the defining constraints of certain region and linearizing the resulting non-linear terms by introducing additional variables and constraints.

Chapter 8

Computational tests

In this chapter, we describe the application of the model to a real test case. Section (8.1) details the realized software implementation. Section (8.2) presents the developed test cases, highlighting their peculiarities, and the results we want to present after solving the models. Section (8.3) delves into various choices made for the different proposed models, while Section (8.4) explores other modeling decisions that may impact solution performance. The chapter concludes with a presentation of the obtained results, notably discussing *modeling* sensitivity in Section (8.5) and *performance* sensitivity in Section (8.6), together with the improvements obtained with the heuristics and the hybrid CSC in Section (8.7).

8.1 Technology implementation

The proposed bilevel models have all been implemented using the Julia language [Bezanson et al., 2017], version 1.9.4. Julia is an open-source programming language expressly designed for quantitative problems, and its ecosystem has a large and fast-growing set of packages to deal with different mathematical problems.

For modeling the optimization problems, the JuMP package belonging to the Julia ecosystem is used, [Lubin et al., 2023]. In particular, a sub-package of JuMP called BilevelJuMP [Dias Garcia et al., 2023] is used to reformulate the bilevel models into single-level models with different readily available approaches such as `StrongDualityMode`, `ProductMode`, or `BigMMode`.

The optimization solver used is Gurobi, [Gurobi Optimization, LLC, 2023] version 11.0, which is natively able to handle bilinear terms such as those found in the revenue adequacy constraints and in some of the proposed objective functions.

For the MILP linearized versions of the models with no bilinear components, we also tested CPLEX version 22.10, [Cplex IBM, ILOG, 2022] Xpress version 9.2.5, [FICO, 2023] and COPT version 7.1.1 [Dongdong et al., 2022] commercial solvers, observing worse computational results than Gurobi, so the main results will be all related to this last solver. In Tab. 8.4 we will report a brief comparison of other commercial solvers.

As already mentioned, the Gurobi solver can handle bilinear constraints internally [Achterberg, 2019]. To do that it also applies *spatial Branch-and-Bound algorithms*, that solve the relaxation of the model obtained by building the McCormick envelopes. Moreover in [Achterberg, 2019] it is proposed to add different types of other cuts. By iteratively partitioning the domain of continuous variables, tighter relaxations can be achieved to prune infeasible solutions (one McCormick polyhedron is turned into two smaller McCormick polyhedra after branching). This means that the bounds of the McCormick relaxation change during the solving algorithm (local bounds).

The PC used for the computational tests is based on a AMD Ryzen 9 7950X CPU with 16 cores and 32 threads, 32 GB RAM DDR5 type running Windows 10 OS.

8.2 Test cases development and output structure

With the support of MBS Consulting Company, a comprehensive Italian, real-size, test case has been defined, which considered one, two, and five representative days. The data were derived from proprietary scenarios subject to non-disclosure agreements based on the objectives of the National Integrated Plan for Energy and Climate (PNIEC) and the Fit for 55 targets. We also defined a small synthetic test case with one zone only, which is useful for testing real optimality and some modeling options for performance issues. The real-size Italian test case is a database of several data visioned to the year 2040 *circa*. It has been developed with the following main peculiarities and educated considerations:

- starting from a certain set of existing power plants of different technologies derived from the installed capacity at the year 2022, several additional capacities per zone and each selected technology are made available for the model to be selected. We focus on fossil-fueled technology[1] such as Combined Cycle Gas Turbine (CCGT), (Open Cycle) Gas Turbine (OCGT) and RES technology such as Photovoltaic, Onshore Wind, Offshore Wind. Additionally, Storage with specific requirements is also envisioned;
- notably for the storage, at the time of writing (December 2023), the EC accepted under EU State aid rules the Italian proposal for long-term auctions in order to install a significant amount of storage[2]. However, the quantity per zone is still unknown[3] and it will be

[1]The existing coal-fired plants are not considered because their phase-out will be completed by the target year 2040

[2]The so called *meccanismo di approvvigionamento di capacità di stoccaggio elettrico* (MACSE) ex Art. 18 of the Legislative Decree 210/21 and regulated by ARERA Decision n. 247/2023/R/eel

[3]An estimate of 9 GW in power and 71 GWh/year in energy has been proposed

precisely defined by Terna later in 2024. Therefore starting from an as-is situation, where almost nothing is available, we estimate a minimum and maximum power to be installed for the storage;

- for each possible future installation with different technologies, investment costs, operative costs and useful life length have been estimated;

- the five representative days are selected from the MBS scenarios by considering the day of 2040 with the expected highest load and one representative day for each season; the two representative days are selected by considering the day of 2040 with the expected highest load and the representative day for autumn, which is the season with the expected lowest load;

- hourly load *zonal* curves are defined for each day(s);

- each RES technology has a capacity factor (i.e. estimate hourly profile of production) for each day(s);

- we do not consider hydropower production explicitly within the GEP model since no additional investments are possible in the Italian system. However, we do take into account their contributions in a simplified manner, i.e. by subtracting from the load zonal profile a typical hydro hourly production profile and adding to it a typical hourly consumption profile for pumping w.r.t the hydro installed in the Italian system using historical aggregated data;

- the electricity market(s) MGP-MI is considered as in the Italian cases with the current several zones, therefore a set of equivalent network constraints were imposed accordingly with an estimate of their future maximum capacity. At the present time, it is unknown if zones will be changed or removed from the Italian market design in the future; this bidding zone review is however a competence of the Italian Authority;

- we set minimum (and maximum) RES penetration level in terms of installed capacity, in accordance with the EC policy targets as declined in the newly modified PNIEC of the Italian Government of July 2023. These minimum and/or maximum levels can be imposed as constraints in the model forcing the model to respect them;

- we set minimum (and maximum) installation level for batteries;

- we set a maximum conventional (CCGT/GT) future level in terms of installed capacity, also estimating the effect of the present Italian CRM and its possible evolution. These maximum levels can be imposed as constraints in the model forcing the model to respect them;

- all thermal offers (CCGT and GT) are set accordingly with their estimated marginal cost, i.e. we simulate a perfect competitive market with no strategic bid-up. This choice has been also made considering that the proposed model is a simil-capacity one with possible side payments;
- as motivated in (4.1.4), we assume that all RES technologies offer their energy at their estimated LCOE;
- at the beginning of each representative day, all thermal units are assumed to be off-line ($\gamma^I_{k,t} = 0$) to account for the maximum start-up costs that these generators could incur in ensuring their revenue adequacy. Additionally, all batteries are assumed to be fully discharged ($B^I_{z,t} = 0$);
- a level of reserve on top of the estimated load has been simply defined as a percentage of the load itself, in order to obtain reserve-proof production systems in terms of adequacy as explained in (5.1.5).

Given the data sets described above, a set of results has been extracted from the solutions of the models and post-aggregated within the Julia code. These are as follows:

1. **For each zone z and for each hour t in the Italian market:**

   ZonRes.1 Net Import/Export from zone, $NetF_{z,t}$

   ZonRes.2 Thermal Accepted Production, $AQ^k_{z,t}$

   ZonRes.3 Photovoltaic Accepted Production, $AQ^v_{z,t}$

   ZonRes.4 Inshore Wind Accepted Production, $AQ^{wi}_{z,t}$

   ZonRes.5 Offshore Wind Accepted Production, $AQ^{wo}_{z,t}$

   ZonRes.6 Storage Charge Quantity, $B^C_{z,t}$

   ZonRes.7 Storage Discharge Production, $B^D_{z,t}$

   ZonRes.8 Required Reserve, $R^{Re}_{z,t}$

   ZonRes.9 Obtained Reserve, $R^{Ob}_{z,t}$

   ZonRes.10 Zonal Price, $\lambda_{z,t}$

2. **For the whole system and for each hour t Aggregated values[4]:**

[4] a unique buying price could be obtained ex-post as an average of the zonal prices $\lambda_{z,t}$ weighed on the total accepted quantities in zone z

AggRes.1 Thermal Accepted Production, AQ_t^k

AggRes.2 Photovoltaic Accepted Production, AQ_t^v

AggRes.3 Inshore Wind Accepted Production, AQ_t^{wi}

AggRes.4 Offshore Wind Accepted Production, AQ_t^{wo}

AggRes.5 Storage Charge Quantity, B_t^C

AggRes.6 Storage Discharge Production, B_t^D

AggRes.7 Required Reserve, R_t^{Re}

AggRes.8 Obtained Reserve, R_t^{Ob}

AggRes.9 Number of new generators for each technology and number of new batteries

3. **For the whole system and for each technology**

Agg$^{\pi}$.1 π^{TOT}

Agg$^{\pi}$.2 u

8.3 Modeling Sensitivity

The several quantities in the output may vary depending on the different conceptual choices made in terms of modeling. Keeping fixed the data set, we therefore performed a modeling sensitivity primarily w.r.t the GEP-RA (1) and GEP-RA (2) model described in Section (4.2) and subsequently detailed in the whole Chapter (5).

By construction, we have:

- For GEP-RA (1) we keep fixed modeling decisions related to:

M1F.1 `OF Model`, (5.3)

M1F.2 `UseRA for Thermal`, (5.31)

- For GEP-RA (2) we keep fixed modeling decisions related to:

M2F.1 `OF Model`, (5.45)

M2F.2 `UseRA for Thermal`, (5.31)

M2F.3 `UseRA for Photovoltaic`, (5.56)

M2F.4 `UseRA for Inshore Wind`, (5.57)

M2F.5 `UseRA for Offshore Wind`, (5.58)

M2F.6 `UseRA for Battery`, (5.59)

A user can also discard for GEP-RA (1) the RA for thermal and some RA for a specific technology, e.g. Offshore Wind, in the GEP-RA (2).

These are additional modeling parameters are valid for both GEP-RA (1) and GEP-RA (2):

Add.1 `Avoid Physical Withholding`, (5.11)-(5.13) $\in \{Y,N\}$

Add.2 `Use Thermal UB`, relaxing (5.7) $\in \{Y,N\}$

Add.3 `Use RES LB and UB`, relaxing one of the constraints in (5.4) and (5.5) $\in \{Y,N\}$

Add.4 `Use Battery LB and UB`, relaxing one of the constraints in (5.6) $\in \{Y,N\}$

Add.5 `UseReserve`, relaxing (5.24) $\in \{Y,N\}$

Add.6 `HowToReserve`, if `UseReserve == Y`, setting (5.24) or (5.43)

8.4 Performances Sensitivity

Given the data set, and a set of modeling choices as defined in (8.3), another sensitivity dimension is related to performance sensitivity. Indeed the complete bilevel model appears very complex to solve even for the best-chosen solver, Gurobi. We compared specific choices combining different parameters as defined below:

P.1 `UseOF linearized` for GEP-RA (2) only, (5.45) as derived in (6.2) $\in \{Y,N\}$

P.2 `UseRA linearized` for all RA considered as derived in (6.2) $\in \{Y,N\}$

P.3 `AddBound dual` using values defined in Tab. 6.1 $\in \{Y,N\}$

P.4 `Extended formulation for price setting` as derived in (5.3) $\in \{Y,N\}$

P.5 `Monolithic Approach` $\in \{Y,N\}$

P.6 `Non-default solver's parameters` $\in \{Y,N\}$

Using the automatic reformulation of BilevelJuMP we observed a clear superiority of the SOS1 reformulation, `SOS1Mode`, when compared to alternative reformulation approaches such as `StrongDualityMode`, `ProductMode`, or `BigMMode`.

Using the monolithic approach (P.5 `== Y`) of the software implementation, therefore using only JuMP and not BilevelJuMP, we can reformulate the Lower Level CSC in different ways as described in (6.1). In particular, for the Fortuny-Amat and McCarl reformulation, with the manually derived single-level model we are able to set tighter values of big-M, as defined in Tab. 6.1 and Tab. 6.2, than the general values (one for primary variables and constraints and one for dual ones) allowed by the automatic reformulation. In this way, we are also able to implement a two steps mode, that as step one envisions a relaxation of the CSC, i.e. an High-Point Relaxation, see Def. (3.2.3), and as step two uses the complete, or partial solution, of this relaxation as a warm start for the complete model[5]. Additionally, using only JuMP with this monolithic version enables other possibilities such as imposing as lazy constraints some set of linear constraints or setting custom branching priorities, see [Gurobi Optimization, LLC, 2023] and [Lubin et al., 2023]. These advanced settings are not possible at the moment using BilevelJuMP.

8.5 Modeling Results and discussion

The models that we compared are

1. Complete GEP-RA (1) as defined in (M1F.1),(M1F.2)
2. Complete GEP-RA (2) as defined in (M2F.1)-(M2F.6)
3. GEP-RA (1) without RA, referred to as GEP (1), as defined in (M1F.1)
4. GEP-RA (2) without RA, referred to as GEP (2), as defined in (M2F.1)

including as Y all the additional parameters (Add.1)-(Add.6).

We report the results obtained with *one* representative day of 2040. It is important to note that this representative day is a summer-time day, with the highest forecast consumption to ensure that the system is adequately designed to meet peak demand. The cardinality of each set is given in Tab. 8.1. The MIQCP GEP-RA (2) model with the Product formulation (6.1) for CSC results in 9,591 (2,647 integer, 2,664 binary, 4,280 continuous) variables and 15,940 (3,464 quadratic, 12,476 linear) constraints, with 41,864 non-zeros.

$\lvert T \rvert$	24
$\lvert Z \rvert$	7
$\lvert K \rvert$	27
$\lvert L \rvert$	6
$\lvert I \rvert$	10
$\lvert Q \rvert$	7

Table 8.1: Cardinality of the main sets of the model considering one representative day

[5] Gurobi accepts also *partial* MILP solution as a warm start and tries to recover a feasible solution

Tab. 8.2 provides a comparison of objective function values and possible economic losses for each model. The economic losses are calculated after the optimal solutions have been obtained, and, when combined with the objective function values, allow the total cost to be calculated. For models (1), the total cost encompasses investment costs, operational costs, and side payments — both those included in the OF value and those computed ex-post; in particular, losses in GEP-RA (1) refer to all technologies except thermal, as its RA is guaranteed within the model. On the other hand, for models (2), the total cost represents the expenses incurred by consumers, combining market costs and side payments; in GEP-RA(2), losses are zero, as RA is guaranteed within the model for all technologies.

For models with RA, we did not achieve a zero optimality gap within the imposed time limit (7,200s); therefore, we present the best results in terms of primal objective values, among those in Tab. 8.5.

[€]	**GEP (1)**	**GEP-RA (1)**	[€]	**GEP (2)**	**GEP-RA (2)**
Costs Inv+Op(+u)	1.104E+08	1.115E+08	Costs_system(+u)	1.342E+08	1.483E+08
Losses	2.876E+06	8.913E+05	Losses	1.901E+07	0
TOT	**1.133E+08**	**1.124E+08**	**TOT**	**1.532E+08**	**1.483E+08**

Table 8.2: Comparison of total costs, given by the sum of objective function values and economic losses. Losses are calculated ex-post for those technologies whose revenue adequacy is not guaranteed within the model.

A comparison of the models in terms of investment decisions is presented in Fig. 8.1. The left-hand graph illustrates the variation in the number of new RES generators and batteries (each with a capacity of 10 MW) between GEP (1) - that can be considered the “traditional” model due to the absence of RA constraints and the minimization of investment and operational costs - and the other models incorporating RA constraints or different objective functions. In GEP-RA (1), the consideration of the RA for thermal plants leads to investing in a smaller number of photovoltaic units, which are those presenting offers at the lowest price. In GEP (2), the rethinking of the objective function, aimed at minimizing consumer costs, results in larger investments in solar, offshore wind, and especially batteries. Conversely, in GEP-RA(2), where the RA is guaranteed for all technologies, there is a notable decrease in offshore wind and an increase in inshore wind. The right-hand graph shows the total installed capacity, including thermal capacity and the capacity of both the existing and the new generators. The thermal capacity is reduced in the model GEP-RA (2).

In Fig. 8.2, we compare the aggregated values (AggRes.1)-(AggRes.8) for different models, illustrating the typical M-shaped curves representing demand and its fulfillment by accepted

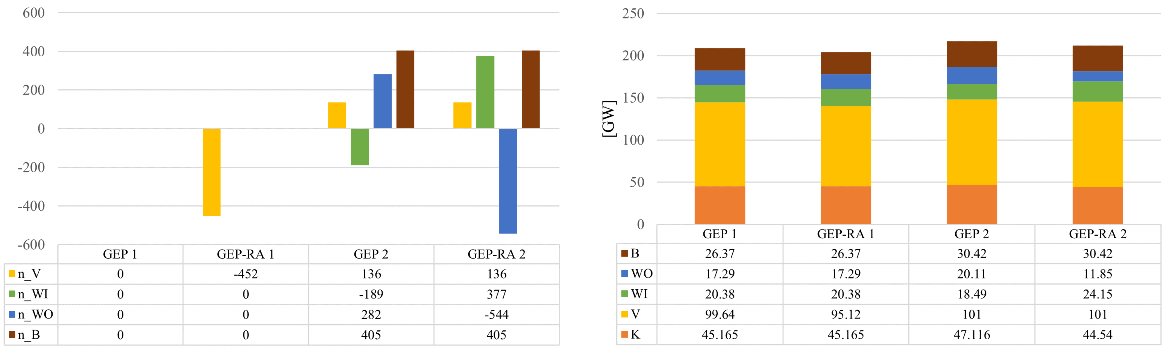
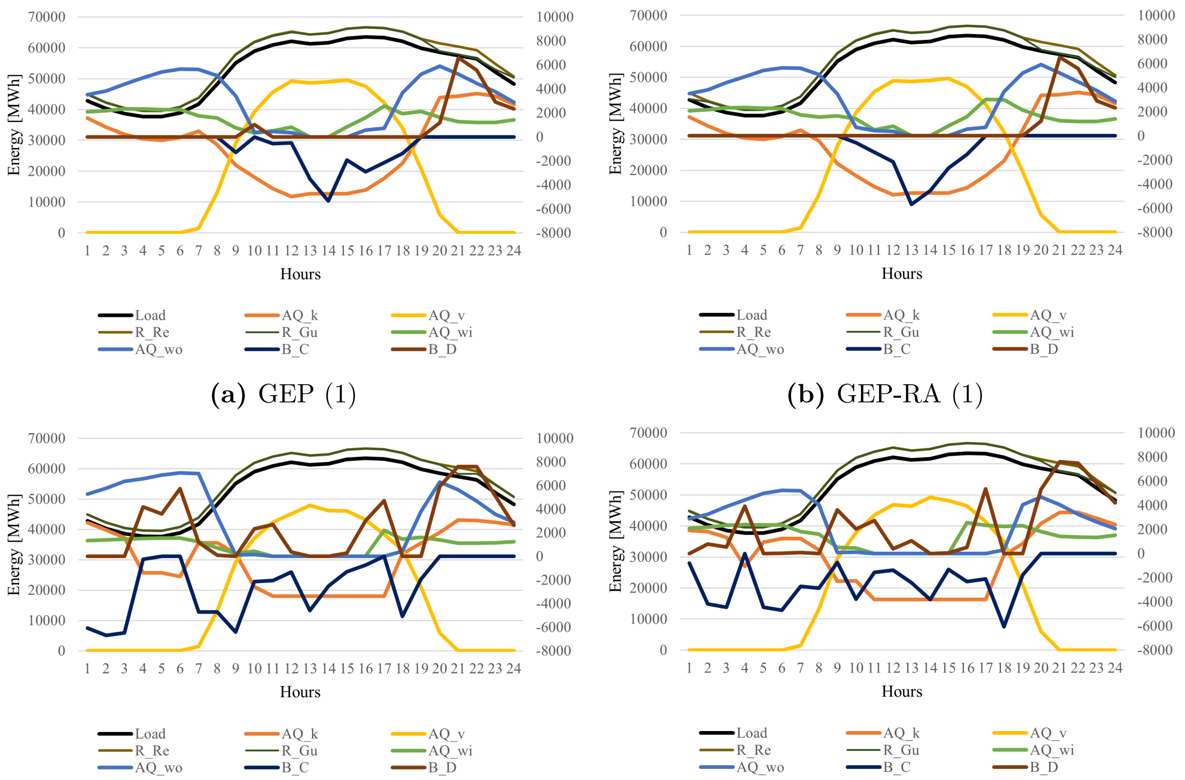


Figure 8.1: Comparison of changes in the number of new RES and batteries between models, with GEP (1) as baseline (left) and total installed capacity between models (right); both are derived from (AggRes.9)

quantities for different technologies. The required reserve is often obtained but in some hours, for the fixed value that we choose for the penalty. The profiles of the technologies involved in meeting the demand vary from hour to hour: in all four models, the accepted quantity of inshore and offshore wind decreases when solar energy is available; the accepted quantity of photovoltaic offers follows a typical bell-shaped curve, except for the central hours, where excess energy is not wasted but used to charge batteries. This charging is not included in the accepted quantity for models (1) (8.2a), (8.2b), as they do not account for the charge in the lower-level balance constraints. On the other hand, models (2) (8.2c), (8.2d), include the charge in the accepted quantities. In models (1), charging occurs mainly during the central hours of the day when solar energy is available. Discharging takes place during the nighttime hours when prices are higher, as shown in Tab. (8.3). Models (2) exhibit a higher level of battery operation overall.

Figures 8.3, 8.4, and 8.5 present a detailed breakdown of the previous results based on the market zones NORD, SUD, and SICI, respectively. They allow us to check that charge and discharge are mutually exclusive since we model one battery for each zone, as correctly required by the models. These figures also include additional information such as zonal prices and net network flows, calculated as the difference between exports and imports. Specifically, NORD - the zone with the higher demand - operates as an importing zone, SUD functions as an exporting zone, and SICI primarily engages in exports. This distinction in roles provides valuable insights into the dynamics of each zone and their contributions to the overall results. For example, the zone SICI imports only in some hours in the models without RA. The thermal accepted quantity in the zone NORD exhibits an inverse correlation compared to imports. In the two other zones, instead, the exports are mainly influenced by solar availability in the central hours and by wind in the other hours. The primal contribution of inshore wind is evident in the SICI zone, while offshore wind gains significance in the SUD zone during hours when photovoltaic sources are



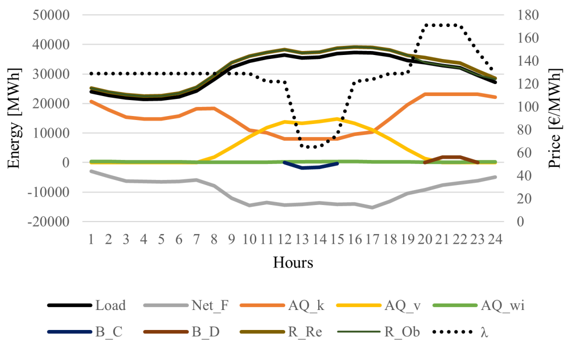
(a) GEP (1) (b) GEP-RA (1)

(c) GEP (2) (d) GEP-RA (2)

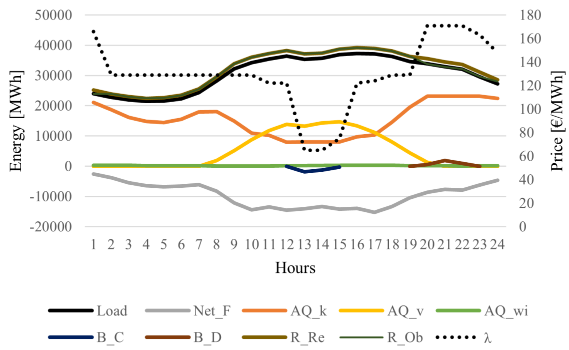
Figure 8.2: (AggRes.1)-(AggRes.8) for each model: the demands, reserves (AggRes.7, AggRes.8) and accepted quantities for thermal (AggRes.1) and photovoltaic (AggRes.2) are plotted with respect to the vertical axis on the left, while the accepted quantities for wind (AggRes.3, AggRes.4) and battery operations (AggRes.5, AggRes.6) are plotted with respect to the vertical axis on the right

unavailable. Thermal generators set the prices in the nighttime hours allowing the discharge to be profitable. In the NORD zone, batteries efficiently charge during periods of lower zonal prices and discharge when prices rise. The battery capacity is sufficient to store excess energy effectively. However, in the other two zones, there are losses of energy (the photovoltaic offered quantity is not fully accepted) as the capacity of lines is not sufficient to export it, and thus the accepted quantity follows the export. This gives a clear indication that the considered transmission capacity could be enlarged. In models (1), where the battery investment cost is in the OF, the energy losses may also be caused by the model's decision to build a number of batteries that is inadequate to store all the excess energy as the reduction in costs achievable by load shifting does not justify the investment costs.

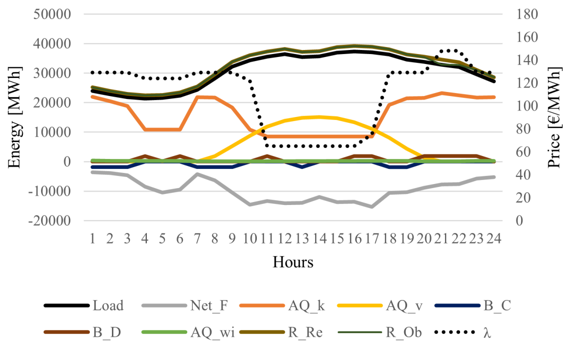
A detailed comparison of the zonal prices is given in Tab. 8.3. The zones are ordered according to the equivalent transmission lines shown in Fig. 4.4, highlighting line congestion when prices between two connected zones differ. For example, in the GEP (1) model during hour 1, the SARD>CSUD line experiences congestion in the opposite direction to the convention:



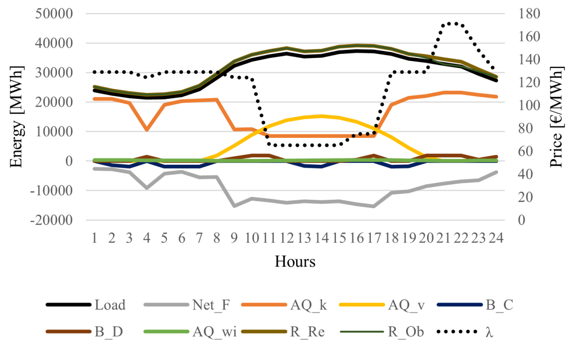
(a) GEP (1)



(b) GEP-RA (1)

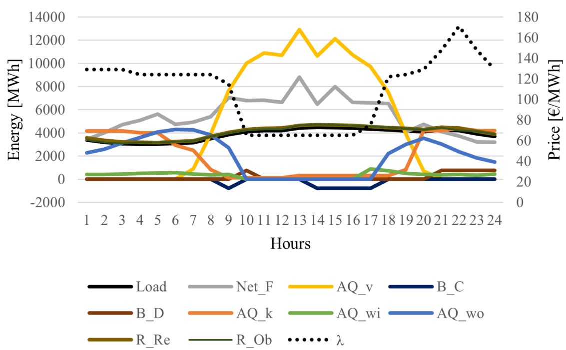


(c) GEP (2)

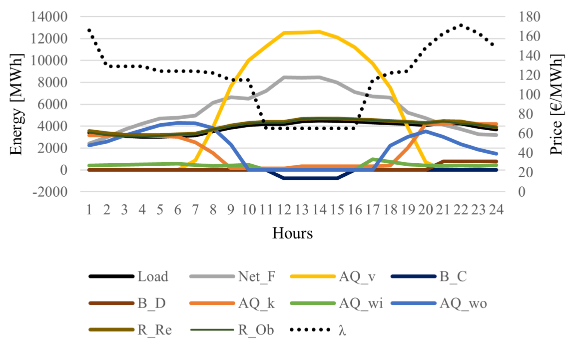


(d) GEP-RA (2)

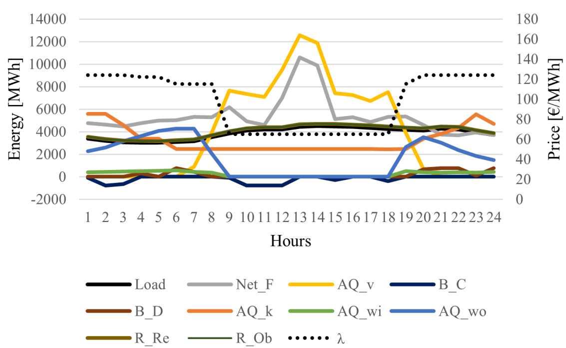
Figure 8.3: (ZonRes.1)-(ZonRes.10) in relation to the zone NORD for each model: only the price (ZonRes.10) is plotted with respect to the vertical axis on the right while all other outputs are plotted with respect to the vertical axis on the left



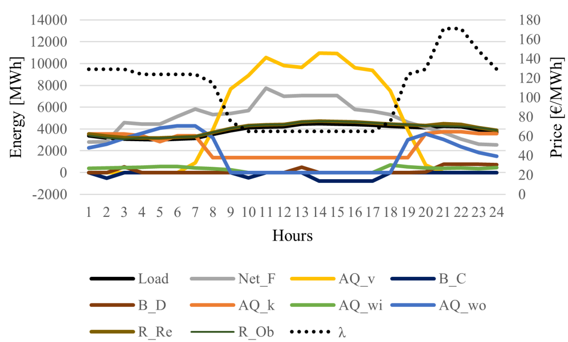
(a) GEP (1)



(b) GEP-RA (1)

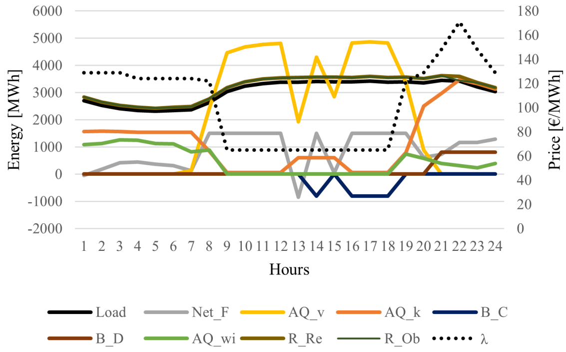


(c) GEP (2)

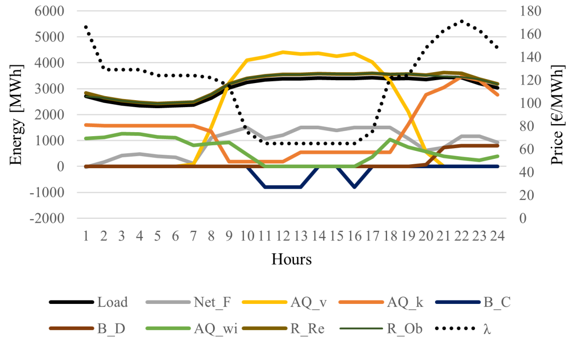


(d) GEP-RA (2)

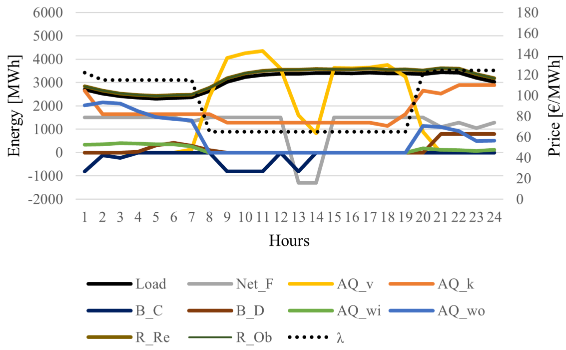
Figure 8.4: (ZonRes.1)-(ZonRes.10) in relation to the zone SUD for each model



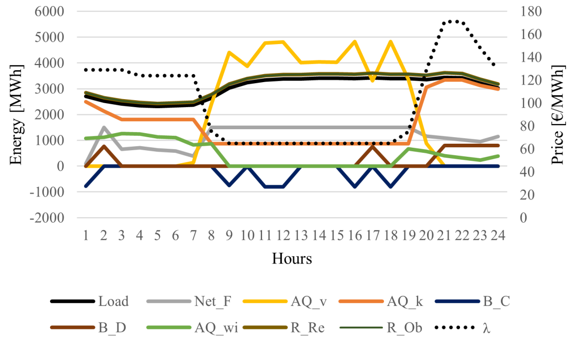
(a) GEP (1)



(b) GEP-RA (1)



(c) GEP (2)



(d) GEP-RA (2)

Figure 8.5: (ZonRes.1)-(ZonRes.10) in relation to the zone SICI for each model

the flow saturates the capacity of the line from CSUD to SARD ($F_{l,t} = \underline{F}_l$), resulting in a higher price in the import zone (SARD). Congestion is also present in some cases even if the price is the same since we consider the same offer price for photovoltaic in different zones. Models incorporating RA exhibit slightly higher prices compared to those without RA, while still maintaining the interpretability of prices. Conversely to the model of [Guo et al., 2022], where the RA is fictitiously guaranteed by artificial price spikes, that cannot be replicated in reality, in our model the price is set by the offer of the marginal generator.

The national single price, PUN, is in this context ex-post calculated as the weighted average of zonal clearing prices for each hour, considering the respective accepted quantities, also because it is known that at some point in 2024 will be removed from the ex-ante calculation within the clearing model that requires a non-convex equality constraint.

Fig. 8.6 shows total profits (Agg$^{\pi}$.1) and side payments (Agg$^{\pi}$.2) for different models, as the sum on zones for a given technology. If a specific model presents a negative total profit also for a single technology, that model is not worth considering.

Photovoltaic and inshore wind are profitable in all four models, even if their RA is not enforced by constraints. Batteries, on the contrary, are always loss-making unless the model guarantees their RA through side payments, that lead them to break even; note that in the test

GEP (1)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
NORD	129	129	129	129	129	129	129	129	129	129	122	122	65	65	75	122	124	129	129	171	171	171	148	129
CNOR	129	129	129	129	129	129	129	129	129	129	122	122	65	65	75	122	124	129	129	171	171	171	148	129
CSUD	129	129	129	129	129	129	129	129	129	129	122	122	65	65	75	122	124	129	129	171	171	171	148	129
SARD	148	129	129	129	129	129	129	129	115	65	65	65	65	65	65	65	65	65	115	171	171	171	148	148
SUD	129	129	129	124	124	124	124	124	115	65	65	65	65	65	65	65	75	122	124	129	148	171	148	129
CALA	129	129	129	124	124	124	124	124	115	65	65	65	65	65	65	65	75	122	124	129	148	171	148	129
SICI	129	129	129	124	124	124	124	122	65	65	65	65	65	65	65	65	65	65	122	129	148	171	148	129
PUN	129.16	129	129	127	127.01	127.05	127.17	127.17	119.98	108.11	103.15	103.44	65	65	71.74	103.59	106.98	119.80	126.74	158.22	164.03	171	148	129.20

GEP-RA (1)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
NORD	166	129	129	129	129	129	129	129	129	129	122	122	65	65	75	122	124	129	129	171	171	171	163	148
CNOR	166	129	129	129	129	129	129	129	129	129	122	122	65	65	75	122	124	129	129	171	171	171	163	148
CSUD	166	129	129	129	129	129	129	129	129	129	122	122	65	65	75	122	124	129	129	171	171	171	163	148
SARD	166	129	129	129	129	129	129	129	115	65	65	65	65	65	65	65	65	65	115	171	171	171	163	148
SUD	166	129	129	129	124	124	124	122	115	115	65	65	65	65	65	65	115	122	124	148	163	171	163	148
CALA	166	129	129	129	124	124	124	122	115	115	65	65	65	65	65	65	115	122	124	148	163	171	163	148
SICI	166	129	129	129	124	124	124	122	115	75	65	65	65	65	65	65	75	122	124	148	163	171	163	148
PUN	166	129	129	129	127	127.05	127.17	126.68	124.09	118.85	103.15	103.44	65	65	71.74	103.59	115.92	124.29	126.90	164.02	168.57	171	163	148

GEP (2)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
NORD	129	129	129	124	124	124	129	129	129	122	65	65	65	65	65	65	75	129	129	129	148	148	129	129
CNOR	129	129	129	124	124	124	129	129	129	122	65	65	65	65	65	65	75	129	129	129	148	148	129	129
CSUD	129	129	129	124	124	124	129	129	129	122	65	65	65	65	65	65	75	129	129	129	148	148	129	129
SARD	129	129	129	124	124	124	129	129	65	65	65	65	65	65	65	65	65	65	65	129	148	148	129	129
SUD	124	124	124	122	122	115	115	115	65	65	65	65	65	65	65	65	65	65	115	124	124	124	124	124
CALA	124	124	124	122	122	115	115	115	65	65	65	65	65	65	65	65	65	65	115	124	124	124	124	124
SICI	122	115	115	115	115	115	115	65	65	65	65	65	65	65	65	65	65	65	65	122	124	124	124	124
PUN	127	126.39	126.31	122.39	122.44	120.38	124.65	120.97	107.12	100.59	65	65	65	65	65	65	71.51	109.29	118.24	127.22	140.58	140.46	127.36	127.44

GEP-RA (2)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
NORD	129	129	129	124	124	124	124	75	65	65	65	65	65	65	65	65	65	65	75	129	171	171	148	129
CNOR	129	129	129	124	129	129	129	129	124	124	65	65	65	65	65	75	75	129	129	129	171	171	148	129
CSUD	129	129	129	124	124	124	124	115	75	65	65	65	65	65	65	65	65	75	124	129	171	171	148	129
SARD	129	129	129	124	129	129	129	129	124	124	65	65	65	65	65	75	75	129	129	129	171	171	148	129
SUD	148	148	129	124	129	129	129	129	65	65	65	65	65	65	65	65	65	65	65	129	171	171	148	129
CALA	129	129	129	124	124	124	124	115	75	65	65	65	65	65	65	65	65	75	124	129	171	171	148	129
SICI	129	129	129	124	129	129	129	129	124	124	65	65	65	65	65	75	75	129	129	129	171	171	148	129
PUN	129.17	129.15	129	124	127.22	127.26	127.28	121.38	103.74	103.43	65	65	65	65	65	71.50	71.37	110.52	120.86	129	171	171	148	129

	λ_k
V	65
WI	75
WO	115
K1	122
K2	124
K3	129
K4	148
K5	163
K6	166
K7	171

k	P_k
n_5_CCGT	122
n_7_GT	122
cs_2_CCGT	122
s_2_CCGT	122
s_5_CCGT	122
cal_1_CCGT	122
sic_1_CCGT	122
sic_2_CCGT	122
cs_3_CCGT	124
s_1_CCGT	124
cal_3_CCGT	124
n_3_CCGT	129
n_8_CCGT	129
cn_2_CCGT	129
sic_5_CCGT	129
n_6_CCGT	148
cn_1_GT	148
cn_4_CCGT	148
cs_5_CCGT	148
sar_1_CCGT	148
sar_2_GT	148
sic_6_GT	148
s_3_GT	163
cs_1_GT	166
cn_3_GT	171
cal_2_GT	171
sic_4_GT	171

Table 8.3: Zonal market clearing prices (ZonRes.10) for each model: in columns are the 24 hours and in rows are the zones, followed by the PUN computed ex-post. The colors indicate which technology sets the price: Photovoltaic (yellow), Inshore Wind (green), Offshore Wind (blue), Thermal (orange). Shades of orange represent the offer prices of the thermal power plants, with darker shades indicating higher prices. The legend on the right lists the offer price for each of the 27 thermal plant types considered.

case considered we do not ask for a return on investment but only for cost recovery. Offshore wind incurs losses in the GEP(1) and GEP(2) models but achieves positive profits in the GEP-RA(1) model, even though its RA is not enforced. In the GEP-RA(2) model, total offshore profits are non-zero, but side payments are positive because, despite considering the same offshore investment costs in different zones, market profits are positive in some zones (not requiring side payments) while negative in others (requiring side payments). In particular, the offshore wind generators in the CALA and SUD zones are profitable within the market (125.5 k€ and 391.1 k€), while those in the CSUD zone incur market losses of 335.6 k€, which are covered by the side payment. Thermal generators make losses in the models without RA, especially the Combined Cycle Gas Turbine (CCGT). Conversely, in the models with RA, they either make a market profit or break even thanks to the side payments. In particular, in GEP-RA (1), the total thermal profits are positive, indicating that some generators are profitable in the market, whereas in GEP-RA (2), all generators with new investments experience losses in the market. It is worth noting that photovoltaic generators enjoy quite high profits in all the models.

The model GEP-RA (2) considers side payments for those technologies that incur losses on

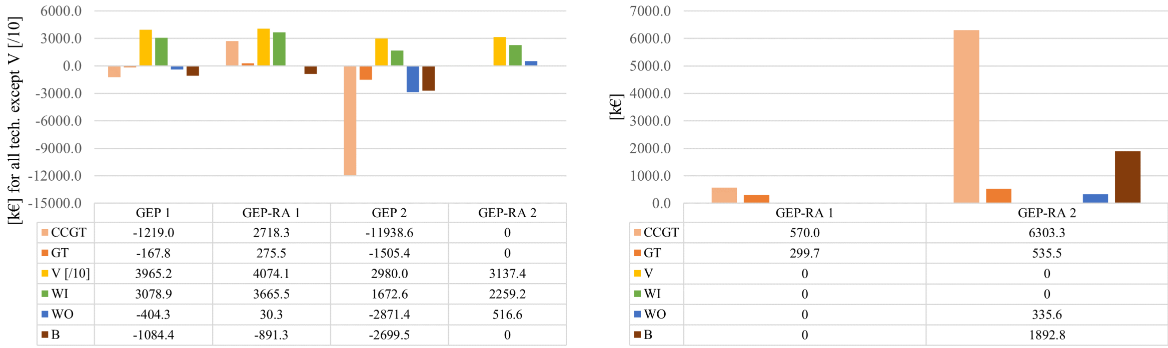


Figure 8.6: Comparison of total profits (Agg$^{\pi}$.1) and side payments (Agg$^{\pi}$.2) among different models; side payments refer only to models with RA: GEP-RA(1) guarantees RA exclusively for thermal power plants, while GEP-RA(2) guarantees RA for all technologies

the market while getting the market clearing prices. One may also hypothesize a two-ways CfD, with a strike price that would redistribute the economic matches and thus possibly increase the side payments required for RA.

8.6 Performances Results and discussion

Tab. 8.4 shows the performance superiority of Gurobi over other solvers. The comparison is made on the MIP version of the GEP-RA model (2), with OF and RA linearized (P.1,P.2 `== Y`) and CSC reformulated using the big-M approach. No additional parameters are taken into account (P.6 `== N`). Xpress and CPLEX produce suboptimal solutions, while COPT fails to find a feasible solution within the given time limit, even with the non-default parameter `HeurLevel` set to 3, imposing an aggressive use of heuristics. It is worth mentioning that the warm start solution obtained by Algorithm (7) is accepted by COPT, which however is not able to improve the initial solution provided, working only on the best bound and showing a gap of 20.50% after 7,200 seconds.

Solver	Gap [%]	Time2FSol	GapFSol [%]	Expl. nodes
GUROBI	4.41	312	10.30	28,913
XPRESS	9.01	1,741	12.42	5,271
CPLEX	20.76	3,627	63.07	9,596
COPT	no sol	no sol	no sol	8,021

Table 8.4: Comparison of solvers for the linearized GEP-RA (2) model (P.1,P.2 `== Y`) with a time for the last solution set to `7,200 secs`. Both the dual variable bounds and the monolithic version with the Big-M reformulation for CSC are considered (P.3,P.5 `== Y`), while the extended formulation and non-default parameters are not considered (P.4,P.6 `== N`). The gap for the last solution, both the time and the gap for the first solution, and the number of explored nodes are shown.

Focusing on Gurobi, in Tab. 8.5, the models are compared according to performance criteria (P.1)-(P.6). With reference to (P.6), as default parameters we consider `Nonconvex=2`, which allows Gurobi to handle non-convex quadratic terms, and the stop criteria `TimeLimit=7,200 secs`, `MIPGap=1e-7` controlling the resolution time and the MIP optimality gap. It should be noted that Gurobi is only guaranteed to be deterministic if exactly the same model is solved with exactly the same parameters; even a small change, such as decreasing or increasing the required `MIPGap`, could change the solution path. In fact, Gurobi uses this information in a proprietary way, possibly differentiating the cuts, the heuristics, and many other internal strategies, thus (slightly) altering the entire path solution. As additional parameters, we consider:

- `MIPFocus`, that allows to modify the focus between finding new feasible solutions and proving that the current solution is optimal
- `PreSOS1BigM`, that allows to specify the largest big-M that can be introduced by the internal automatic reformulation of SOS1 constraints into binary form

In particular, we set the `MIPFocus` parameter to 1 due to challenges encountered by the solver when seeking feasible solutions for models involving RA. For models without RA, we exclude the consideration of this parameter as feasible solutions are quickly identified, and the solver can prove their optimality. We set the value of `PreSOS1BigM` to $50,000$, which corresponds to an upper approximation of the maximum big-M calculated for our test case, as derived from the analysis of the upper bounds presented in Tab. 6.1 and Tab. 6.2. The `PreSOS1BigM` parameter is omitted when the reformulation directly entails big-M constraints instead of SOS1 constraints.

For GEP (1) and GEP (2) it is possible to achieve a zero MIP gap in reasonable times. Conversely, GEP-RA (1) and GEP-RA (2) fail to close the MIP gap within the `TimeLimit`. The higher computational cost is justified by the results shown in Tab. 8.2: although the solutions with RA have a MIP gap greater than 0, they have a lower total cost.

For GEP-RA (2), i.e. the most complex model that includes RA constraints for all technologies, Gurobi fails to find a feasible solution within the `TimeLimit` without the activation of performance parameters (P.1-P.6 `== N`), notably with P.1, P.2 `== N` which consider a natural bilinear formulation; instead, the linearization of the model (P.1, P.2 `== Y`) allows to obtain feasible solutions up to a minimum gap of about 4.34%. The inclusion of bounds for dual variables (P.3 `== Y`), as well as additional parameters (P.6 `== Y`), further reduces the gap to 1.76 %, although the time to find the first feasible solution remains high. The monolithic approach (P.5 `== Y`) significantly diminishes this time by about 90%; in particular, the two-step version performs better in terms of the gap. The best gap of 1.65% is achieved by the extended formulation (P.4 `== Y`) in addition to the dual bounds and the non-default parameters, which shows also an improved root relaxation value. The best primal objective value is instead obtained with the linearized version (P.1, P.2 `== Y`), together with the dual bounds and the non-default parameters, but the gap is large due to the worst best bound.

For GEP-RA (1), the best gap of 1.12% is obtained by the extended (P.4 `== Y`) linearized (P.2 `== Y`) version, with bounds for dual variables (P.3 `== Y`) and additional parameters (P.6 `== Y`) thanks to the better best bound, with 60% fewer nodes explored than in other tests. The monolithic version (P.5 `== Y`), in this case as well, has the ability to reduce the time to find the first feasible solution. Moreover, it shows the best primal objective value.

Model	P.1-P.6	Primal	Dual	Time2LSol	Gap [%]	Time2F(S)Sol	GapF(S)Sol [%]	Root rel.	Expl. nodes	Note
GEP (1)	NNYNNY	110,441,160	110,441,159	437.64	0	23 (24)	2.91 (2.79)	107,657,300	106,717	
GEP (2)	NNYNNY	134,156,103	134,156,103	617.04	0	56 (56)	5.18 (4.34)	104,657,300	31,326	
GEP-RA (1)	NNNNNN	no sol	108,844,445	7,200	no sol	no sol	no sol	107,867,100	549,031	
GEP-RA (1)	NYNNNN	111,675,650	109,366,384	7,200	2.07	6,134 (6,150)	6.29 (5.54)	107,629,600	791,595	
GEP-RA (1)	NNYNNY	111,552,212	109,565,003	7,200	1.78	153 (157)	68.60 (4.86)	107,910,800	407222	
GEP-RA (1)	NNYNYY	111,552,212	109,529,379	7,200	1.81	127 (129)	8.23 (4.29)	107,910,800	510,007	BIGM 1 solve
GEP-RA (1)	NNYNYY	111,523,200	109,670,328	7,200	1.66	164 (168)	74.80 (5.05)	107,910,800	461,818	BIGM 2 solve
GEP-RA (1)	NYYYNY	111,559,972	110,307,583	7,200	1.12	273 (276)	69.90 (6.34)	107,635,200	157,792	
GEP-RA (1)	NYYNNY	111,567,855	110,208,215	7,200	1.22	286 (289)	70.10 (8.54)	107,635,200	162,096	
GEP-RA (2)	NNNNNN	no sol	135,228,358	7,200	no sol	no sol	no sol	127,962,300	120,371	
GEP-RA (2)	YYNNNN	148,444,155	141,994,964	7,200	4.34	533 (537)	77.10 (32.30)	103,032,500	51,761	
GEP-RA (2)	NNYNNY	148,338,740	145,732,322	7,200	1.76	459 (459)	69.70 (22.40)	129,538,600	40,048	
GEP-RA (2)	NNYNYY	148,293,460	144,688,192	7,200	2.43	35 (48)	78.30 (9.87)	128,718,000	21,120	BIGM 1 solve
GEP-RA (2)	NNYNYY	148,438,159	145,755,327	7,200	1.81	35 (50)	78.40 (9.90)	128,278,000	27,651	BIGM 2 solve
GEP-RA (2)	NNYYNY	148,391,023	145,936,474	7,200	1.65	308(379)	73.50 (41.40)	129,630,200	26,839	
GEP-RA (2)	YYYNNY	148,292,836	142,149,215	7,200	4.14	434 (442)	10.50 (10.40)	103,333,000	45,886	

Table 8.5: Performance results according to (P.1)-(P.6); for each model and combination of performance criteria, the value of the primal objective, the value of the dual objective (best bound), the time to the last solution, the optimality gap, the times to the first and second solutions, the gaps for the first and second solutions, the root relaxation value, and the number of explored nodes are shown.

8.7 Performances Results with heuristics and hybrid CSC

In Section 8.6, the focus was on the SOS1 reformulation for CSC, highlighting its superior performance, especially when using the Gurobi parameter `PreSOS1BigM`. We now turn our attention to presenting the results of using the formulations detailed in Chapter 7. Here the objective is to isolate the impact of the proposed feasibility cuts, warm starts, and, above all, hybridization. Consequently, the comparison of results is conducted without leveraging additional Gurobi parameters (P.6 `== N`).

The analysis in Tab. 8.6 is carried out by comparing *baseline formulations*, which comprise a single CSC formulation, with *hybrid formulations*, which consider different combinations of CSC. The monolithic version of the code is used (P.5 `== Y`) since BilevelJuMP does not support combining multiple CSC reformulations[6].

In addition to the previous parameters (P.1)-(P.6), the following parameters are considered:

P.7 `CSCReformulation` $\in \{1,2,3,4,5\}$

1 Product Formulation (6.1a)-(6.1l)

2 Single CSC (A.3)

3 Big-M (A.5a)-(A.5q)

4 SOS1 (A.4a)-(A.4q)

5 SD (A.6)

P.8 `HybridPM` (6.1a)-(6.1l) as redundant $\in \{Y,N\}$

P.9 `HybridSD` (A.6) as redundant $\in \{Y,N\}$

P.10 `MultipleSD` one SD constraint for each $t \in T \in \{Y,N\}$

P.11 `BoundsPM` as defined in (7.15)[7] $\in \{Y,N\}$

P.12 `AddFeasCut` as defined in (7.2) $\in \{Y,N\}$

P.13 `UseWS` as illustrated in Algorithm (7) $\in \{Y,N\}$

P.14 `UsealternativeWS` as illustrated in Algorithm (8) $\in \{Y,N\}$

P.15 `MyPrimalHeur` callbacks $\in \{Y,N\}$

[6]The Mixed mode of BilevelJuMP allows for the selection of a specific reformulation for each CSC. However, it restricts the choice to only one reformulation per CSC, preventing the use of multiple reformulations for a single CSC simultaneously as we propose here.

[7]utilizing the Ms of Tab. 6.1 and Tab. 6.2

When either `UseWS == Y` or `UsealternativeWS == Y`, each model variable can be warm started by setting the corresponding binary parameter `Start == 1`.

The following additional indicators are also considered, where UB and LB denote the upper and lower bounds of the model at the time or node of evaluation for the run under consideration, while $BestUB$ and $BestLB$ denote the best UB and LB found among different runs:

$$IG_1 = \frac{UB - LB}{UB} \cdot 100 \qquad IG_2 = \frac{UB - LB}{LB} \cdot 100$$
$$PG = \frac{UB - BestLB}{BestLB} \cdot 100 \qquad DG = \frac{BestUB - LB}{LB} \cdot 100$$

The Inherent Gap (either IG_1, as determined by Gurobi, or alternatively IG_2) is specific to each run and does not consider the best UB or LB across different runs. For a more detailed analysis, it is useful to calculate both the Primal Gap (PG) and the Dual Gap (DG), which take into account the best LB and UB, respectively. The PG assesses the quality of the feasible solution, thereby evaluating the efficacy of heuristics and branching, while the DG estimates the ability to prove optimality. In particular, the DG at the root node before and after cuts can be used to assess the tightness of different formulations and the capability of generating useful cuts for a given formulation. We recall that Gurobi performs a branch-and-cut [Gurobi Optimization, 2016]: after the presolve, which tightens the formulation and reduces the problem size, the solver starts from the root node and performs a continuous relaxation that ignores integrality and gives a lower bound (in the case of minimization and referred to as root relaxation value) on the optimal integral objective. The solver then adds cuts to eliminate fractional solutions at the root node, then proceeds along the trees by selecting branching variables and possibly adding more (local) cuts.

The $BestUB$ and $BestLB$ are determined by analyzing the outcomes over both Tab. 8.5 and Tab. 8.6, focusing on identifying the highest lower bound and the lowest upper bound. Notably, the run described in the last row of Tab. 8.5 achieves the best upper bound (the lowest primal objective value is 148,292,836), but falls short of proving optimality (the highest dual value of that table is 145,936,474). In contrast, the results of Tab. 8.6 give a slightly worse upper bound (148,298,344), but show a significant improvement in the lower bound (148,153,221).

Hybridization has been found to be highly effective: its application has resulted in significantly faster primal-dual convergence, as shown in Fig. 8.7. The inherent gaps with single CSC (left) and hybrid CSC (right) are compared in Fig. 8.8.

Among the baseline runs, the one with the product reformulation (P.7 `== 1`) obtains a good LB, as evidenced by the low value of DG; however, it is not able to find a good UB, as evidenced

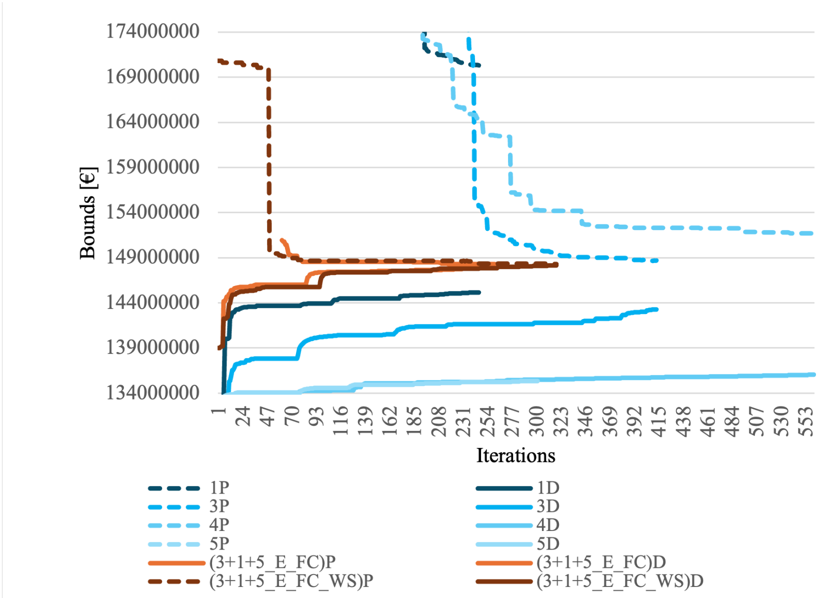


Figure 8.7: Evolution of upper (dashed) and lower (solid) bounds of GEP-RA (2) as a function of solver iterations, given a `TimeLimit` of 7200 secs: single CSC and hybrid CSC formulations are compared. The numbers indicate the CSC formulations: product (1), big-M (3), SOS1 (4), and SD (5). Extended formulation (E), feasibility cut (FC), and warm start (WS) are also considered.

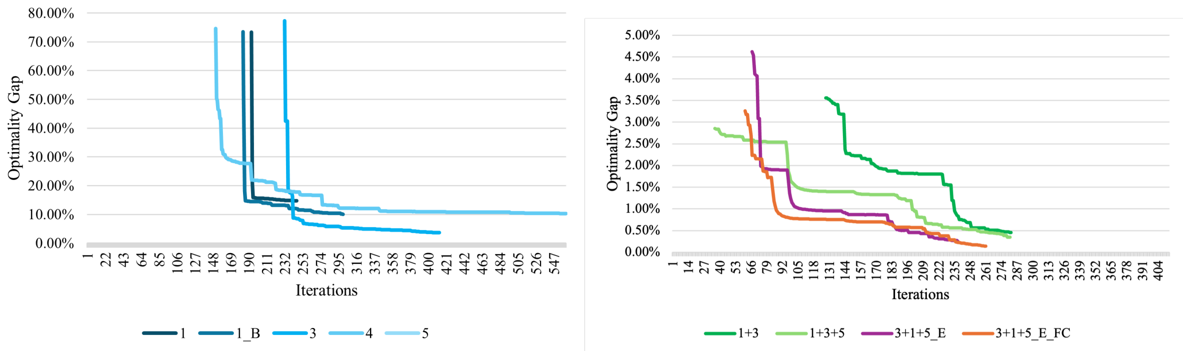


Figure 8.8: Inherent gaps comparison between single CSC (left) and hybrid CSC (right) formulations. The numbers indicate the CSC formulations: product (1), big-M (3), SOS1 (4) and SD (5). The formulation with the upper bounds for each term of the CSC products (B), the extended formulation (E), and the feasibility cut (FC) are also considered.

by the high value of PG. Upon consideration of the M bounds (P.11 `== Y`), there is a slight improvement in the results. The big-M reformulation (P.7 `== 3`) yields the most favorable results, particularly in terms of the UB; the application of either the extended reformulation (P.4 `== Y`) or the feasibility cut (P.12 `== Y`) enhances the LB. The SOS1 reformulation (P.7 `== 4`) exhibits suboptimal performance. Finally, the SD reformulation (P.7 `== 5`), either applying SD for each $t \in T$ or through a singular constraint summing over t, exhibits the least favorable outcome, being unable to obtain a feasible solution within 7,200 secs.

Among the hybrid runs with the combination of two CSC formulations, the one with big-M

(P.7 `== 3`) and product (P.8 `== Y`) shows improved results. The addition of the product to the already effective big-M formulation in terms of good UB helps the solver prove optimality by improving the LB. Enlarging the hybridization approach by combining the three formulations: big-M, product, and SD (P.9 `== Y`) further improves the result, also in terms of the first feasible solution found in a shorter time and with a smaller IG_1. The extended formulation (P.4 `== Y`) and the feasibility cuts (P.12 `== Y`) allow to even further improve the DG values.

The tightness of the hybrid formulations is evidenced by the reduction of the root gap both before cuts ($GDBC$) and after cuts ($GDAC$), together with the decrease in the number of explored nodes.

With one representative day, the warm start (P.13 `== Y`) facilitates an efficient initialization of the branch-and-cut process, although the gap after 7,200 seconds is similar to the one obtained without the warm start. The warm start has been performed with `Start == 1` for $\gamma_{k,t}, n_k^K, n_z^V, n_z^{WI}, n_z^{WO}, n_z^B, \phi_{z,t}, Q_{k,t}^K, Q_{z,t}^V, Q_{z,t}^{WI}, Q_{z,t}^{WO}, F_{l,t}$ but many other combinations are possible. On the other hand, with more than one representative day (two or five days), the solver without a warm start is not able to find a single feasible solution after 12 hours of computation or more, even if general purpose heuristics are applied within Gurobi. When using the proposed algorithms (with either P.13 `== Y` or P.14 `== Y`), the solver instead is able to start with a feasible solution from the beginning and then proceeds with the branch-and-cut path.

Performance Criteria		Last solution						First (Second) solution		Root node	
Combinations	P.4,P.7-P.13	UB	LB	IG1	PG	DG	Expl. nodes	Time	IG1	DG BC	DG AC
1	N1NNNNNN	170,282,471	145,165,098	14.75	14.94	2.06	35,458	3,818 (3913)	73.40 (20.90)	11.52	3.30
1 B	N1NNNYNN	160,717,061	144,498,893	10.09	8.48	2.53	60,720	2,617 (2,669)	73.50 (50.20)	11.52	3.28
3	N3NNNNNN	148,676,639	143,267,741	3.64	0.35	3.41	15,760	1,011 (1,022)	77.30 (42.40)	15.60	7.76
3 E	Y3NNNNNN	148,453,822	145,378,064	2.07	0.20	1.91	23,540	567 (581)	77.30 (9.77)	15.09	7.51
3 FC	N3NNNNYN	149,073,161	143,509,417	3.73	0.62	3.24	17,920	860 (887)	77.30 (7.89)	15.60	7.97
4	N4NNNNNN	151,707,690	136,058,734	10.32	2.40	8.89	127,262	458 (464)	74.70 (50.30)	15.77	12.47
5 (*for* $t \in T$)	N5NNYNNN	no sol	135,333,966	no sol	no sol	9.47	86,332	no sol	no sol	17.71	10.78
5 (*sum* $t \in T$)	N5NNNNNN	no sol	133,878,457	no sol	no sol	10.66	265,177	no sol	no sol	19.04	12.05
3+1	N3YNNNNN	148,430,334	147,748,010	0.46	0.19	0.27	8,128	927 (929)	3.56 (3.56)	10.50	2.13
4+1	N4YNNNNN	148,742,114	146,124,824	1.76	0.40	1.39	15,049	1184 (1223)	65.10 (35.30)	10.80	2.96
3+5	N3NYYNNN	150,817,475	143,514,113	4.84	1.80	3.23	10,746	1995 (2027)	77.00 (36.80)	14.70	6.08
4+5	N4NYYNNN	150,070,038	140,927,488	6.09	1.29	5.13	18,578	2278 (2314)	73.60 (68.00)	15.32	7.48
1+5	N1NYYNNN	150,671,176	144,764,297	3.92	1.70	2.34	41,868	1400 (1417)	73.20 (16.80)	7.23	3.25
3+1+5	N3YYYNNN	148,340,504	147,824,373	0.35	0.13	0.22	7,859	93 (93)	**2.85 (2.84)**	6.71	2.01
3+1 E	Y3YNNNNN	148,343,527	147,985,086	0.24	0.13	0.11	7,024	5 (5)	13.60 (13.50)	10.39	1.79
3+1+5 E	Y3YYYNNN	148,394,455	148,044,114	0.24	0.16	0.07	7,150	605 (605)	4.62 (4.53)	**6.68**	1.83
3+1 E FC	Y3YNNNYN	148,326,042	148,111,030	0.15	0.12	0.03	9,735	6 (6)	13.90 (13.90)	10.39	**1.63**
3+1+5 E FC	Y3YYYNYN	**148,298,344**	148,092,239	0.14	**0.10**	0.04	**5,879**	615 (630)	3.26 (3.17)	**6.68**	1.67
3+1+5 E FC WS	Y3YYYNYY	148,301,624	**148,153,221**	**0.10**	0.10	**0.00**	7,784	**0 (7)**	18.60 (18.60)	**6.68**	1.87

Table 8.6: Performance results according to (P.4),(P.7)-(P.13); each run considers the GEP-RA (2) model, with (P.1),(P.2) `== N`, (P.3),(P.5) `== Y`, (P.6) `== N`, (P.14), (P.15) `== N`, and a time to the last solution set to `7,200 secs`. In blue are runs with single CSC, eventually with bounds (B), extended formulation (E), and feasibility cuts (FC); in green, purple and orange are runs with different combinations of CSC, eventually with E, FC, and warm starts (WS). For each combination, different indicators are shown: for the last solution, the value of the UB (primal objective), the value of the LB (dual objective), the IG, the PG, the DG, and the number of explored nodes; for the first and second solutions, the times and the IGs; for the root node, the DGs before (BC) and after (AC) cuts.

Chapter 9

Future Research Directions

The GEP models that we have proposed in this thesis can be the subject of several future research directions, involving both modeling and formulation as well as algorithmic approaches. In this chapter, we sketch and discuss a few of them.

9.1 Better formulations and specialized heuristics

Some blocks of constraints, and the shape of some objective functions, are difficult to deal with since their inherent nonconvexity. Therefore better formulations of the bilinear components may be identified. In (7.1) we presented a promising heuristic that enhanced the solver's capabilities to find and improve primal solutions. Other specialized heuristics could be developed starting from e.g. some relaxation of the original problem, notably the High-Point Relaxation, see Def. (3.2.3), or the relaxation of the RA coupling constraints such as (5.31) for thermal generators.

9.2 Deepening of the Hybrid CSC approach

Somehow inspired by the approach of [Kleinert, 2021], that in our context did not work well as explained in (7.2), in (7.3) we proposed a novel hybrid CSC formulations for the LL that ensembles different known single reformulations in a reasoned way. In a dedicated section of the computational tests, (8.7), we confirm our intuition by showing enhanced performances on our test instance. Indeed several questions are still open: first of all, a deeper interpretation of the reasons for this better performances, possibly doing an analysis of the (convex hull) set formed by the intersections of the various constraints of the different hybrid CSC formulations. Secondly, it would be of interest to analyze the generality of the approach, even in different contexts. Additionally, we envision possible improvements related to a selective and reasoned inclusion of constraints from each CSC reformulation in order to form different ensembles, possibly lighter in terms of the number of constraints. Lastly, it would be of interest to study the most appropriate reformulations/linearization for those constraints that are bilinear, not to rely solely on the capabilities of the Gurobi solver.

9.3 Decomposition

Decomposition techniques provide an approach for addressing complex problems in a decentralized or distributed manner. They can reduce computational time if the problem under consideration has one of the following structures [Conejo et al., 2006]:

1. primal block angular: the original problem is decomposed by relaxing complicating constraint(s)

2. dual block angular: the original problem is decomposed by fixing complicating variable(s) to given values

In fact, both (1) and (2), make the problem more difficult to solve, preventing a straightforward or block-wise solution. The decomposition approach transforms the problem into a set of subproblems, each of which is easier to solve than the original (non-decomposed) problem.

In our model, the RA constraints, such as (5.31) for thermal generators, link different representative days. In fact, these constraints state that revenue adequacy must be ensured throughout the entire time horizon considered. The *Consensus Alternating Direction Method of Multipliers* (ADMM) can be applied by imposing RA constraints for each representative day and reformulating the problem. This reformulation entails making the investment decision n day-dependent, while ensuring it assumes the same value for each day through the complicating constraint (9.1), where z^{AUX} serves as an auxiliary variable.

$$n_c = z^{AUX},\ c \in \{1, 2, \ldots, C\} \tag{9.1}$$

The ADMM relaxes the constraints that complicate the problem and penalizes their violation in the objective function. Over the iterations, the subproblems defined for each day update the value of n_c, penalizing its difference with respect to the average value among days of the previous iteration; if the penalty is sufficiently large, the optimal solution to the relaxed problem will satisfy the original constraints.

An alternative approach could be the *Variable Splitting* ADMM [Wuijts et al., 2023], where additional variables are introduced as copies of the variable λ, representing the market-clearing price optimized by the lower level and used in the upper level. This separation aims to distinguish the pricing variable used in the upper level from that in the lower level. Throughout the iterations, these two variables are expected to ultimately converge to the same value.

Other general approaches include Lagrangian relaxation, which can be used to obtain lower bounds on the optimal objective value. The key advantage is that when the coupling constraints are dualized, the problem decomposes by blocks. Also, Lagrangian relaxation can provide stronger bounds w.r.t. linear programming (LP) ones.

An alternative, related approach is the Dantzig-Wolfe reformulation [Vanderbeck and Savelsbergh, 2006], which decomposes the problem into a master problem, which may have an exponential number of columns and several subproblems.

9.4 Network inclusion, from GEP to GTEP

On one hand, the GEP assumes a separate process for the electrical network development and of course, this is a simplification. On the other hand Transmission Expansion Planning considers only the grid expansion problem. However, the network is, and must be considered, an electrical asset that facilitates the goals production side, especially if these are strongly driven by policy goals. It is therefore natural to think of integrated models called Generation and Transmission Expansion Planning (GTEP).

Therefore also the proposed models could be enriched by incorporating the planning or reinforcement of the transmission networks. In [Hemmati et al., 2013] the authors survey GEP, TEP, and GTEP approaches, in [Deng and Lv, 2020] is given a perspective on how optimization models have evolved due to the widespread integration of variable renewable energy. The authors identified and analyzed 34 studies related to power system planning that address the growing influence of variable renewable energy.

9.5 Introducing uncertainty

The proposed models could also be enhanced by taking into account long-term uncertainty about, e.g., fuel and CO_2 costs as in [Micheli and Vespucci, 2020], along the load or investments costs avenue. Remaining in the bilevel modeling settings, one could use approaches such as those reviewed in the recent [Beck et al., 2023] where a survey of bilevel modeling under uncertainly is presented. In this work, the authors review classic approaches such as stochastic or robust, e.g. [Bertsimas et al., 2011], techniques.

Chapter 10

Conclusions

This thesis navigates the evolving landscape of global electrical systems, driven by technology, environmental concerns, and changing energy consumption patterns. Focused on one of the most relevant issues of *Generation Expansion Planning*, the research emphasizes the pivotal role of quantitative models, particularly optimization models.

The proposed innovations are manifold, including significant advances in both modeling frameworks and algorithmic strategies. The main modeling contributions are

- Modeling of a bilevel optimization problem that minimizes the cost of the system while ensuring the Revenue Adequacy for different technologies: alternatively to the conventional approach of centralized GEP models, GEP-RA (2) minimizes the total costs for consumers by considering the investment and operational costs incurred by the generators in the RA constraints, rather than in the objective function;

- Optimal solutions in contrast to heuristic solutions, with the aim of achieving the minimum cost of the problem, rather than an approximation calculated ex-post;

- Comprehensive examination of the energy system, encompassing a diverse array of new energy sources and storage options as well as the existing resource mix: the model evaluates the interrelationship between the evolution of battery technology and renewable sources, as well as ensuring revenue adequacy across all technologies;

- Revenue Adequacy with meaningful revenues from the market: zonal hourly prices are determined endogenously and reflect the operations of the Italian electricity market;

- Zonal investments: the incorporation of multiple zones brings to the surface the issue of investment location, particularly relevant in generation expansion, and recently addressed by Italian rulers within the framework of investment adequacy;

- Accounting for start-up costs: the model addresses the non-convexities associated with start-up costs, ensuring a more accurate representation of operational expenses of thermal plants;
- Inclusion of side payments: integration of payments outside of the market as envisaged by the Italian policymakers, with the Capacity Market, the MACSE, and the FERX;
- Innovative offer strategy for RES: solar and wind generators offer at their Levelized Cost of Energy;
- Realistic model decisions: introduction of specific constraints to avoid both strategic withholding and price indeterminacy.

The main algorithmic contributions are

- Exact linearization technique for the revenues: exploits the discretized prices determined by market clearing;
- Warm-Start algorithms: ensure a feasible starting solution at the outset of the solution process;
- Hybrid modeling for CSC: enhances the model's ability to prove the optimality of solutions.

Each of these contributions is explored in depth across the thesis, as detailed below:

- Chapter 1:
  - Acknowledges the ongoing transformation of electrical systems worldwide;
  - Explores the intricacies of electricity markets, emphasizing the central theme of missing money and some remedies given by the Capacity Remuneration Mechanisms, thus paving the way for the concept of revenue adequacy.
- Chapter 2:
  - Conducts a thorough literature review on pricing schemes in diverse markets;
  - Reviews the literature on GEP, distinguishes between traditional GEP and those considering revenue adequacy, addressing non-convexity issues.
- Chapter 3:
  - Introduces elements of Mathematical Programming and presents a general taxonomy for optimization problems;

- Focuses on bilevel optimization problems and their hierarchical structure;
- Discusses reformulation approaches, laying the groundwork for subsequent modeling.

- Chapter 4:
  - Explores modifications to the Italian Market Operator (MO) model;
  - Addresses technical minimum inclusion for thermal power plants in MO;
  - Analyzes the modeling of the GEP Objective Function and introduces two high-level GEP models.
- Chapter 5:
  - Details the two main proposed GEP models, namely GEP-RA (1) and GEP-RA (2), with a formal discussion of the Objective Functions and of all constraints;
  - Introduces innovative formulations to tackle specific issues like withholding strategies;
  - Enhances formulations related to price indeterminacy and presents a mixed-binary version of the models.
- Chapter 6:
  - Derives reformulations of GEP proposed models that can be solved by general-purpose solvers;
  - Explicitly defines optimality conditions of the Lower Level of the resulting bilevel problems;
  - Proposes a simple yet effective linearization of relevant bilinear components in objective functions and constraints.
- Chapter 7:
  - Provides tailored approaches, including feasibility cuts and heuristic algorithms. This significantly improves the solver's ability to find feasible solutions that otherwise cannot be found;
  - Applies cuts proposed by the literature to address the issue of proving optimality and studies why these are ineffective in our context;
  - Introduces an innovative combination of different CSC formulations, in an "hybridized" way, that produces a tighter formulation. This in turn significantly improves the solver's ability in proving the optimality of the solution.

- Chapter 8:
  - Describes the real software implementation made and the construction of an Italian test case;
  - Utilizes the test case for computational tests, comparing proposed models conceptually and in terms of resolution performance;
  - Presents disaggregated and aggregated outputs, discussing simulation results;
  - Highlights the performance improvements obtained with new formulations, in particular the ability to find an initial solution and to narrow the optimality gap.
- Chapter 9:
  - Explores future research directions, including considerations for modeling the contextual transmission network planning and addressing uncertainty;
  - Highlights possible directions for developing more efficient solving methodologies, such as enhanced formulations of specific components and decomposition approaches;
  - Advocates for additional research into the hybrid CSC formulations, focusing on the reasons behind their performance advantages and their versatility across different bilevel problems;
  - Suggests selecting specific constraints for each CSC formulation and exploring advanced linearization techniques.

References

[ACER, 2021a] ACER (2021a). ACER Guidance on the application of Regulation (EU) No 1227/2011 of the European Parliament and of the Council of 25 October 2011 on wholesale energy market integrity and transparency.

[ACER, 2021b] ACER (2021b). Guidance on the application of Regulation (EU) No 1227/2011 of the European Parliament and of the Council of 25 October 2011 on wholesale energy market integrity and transparency.

[Achterberg, 2019] Achterberg, T. (2019). Products of variables in mixed integer programming.

[AGCM, 2021] AGCM (2021). Bollettino settimanale.

[Andrianesis et al., 2022] Andrianesis, P., Bertsimas, D., Caramanis, M. C., and Hogan, W. W. (2022). Computation of convex hull prices in electricity markets with non-convexities using dantzig-wolfe decomposition. *IEEE Transactions on Power Systems*, 37(4):2578–2589.

[Araoz and Jörnsten, 2011] Araoz, V. and Jörnsten, K. (2011). Semi-lagrangean approach for price discovery in markets with non-convexities. *European Journal of Operational Research*, 214(2):411–417.

[Bacci et al., 2023] Bacci, T., Frangioni, A., Gentile, C., and Tavlaridis-Gyparakis, K. (2023). New mixed-integer nonlinear programming formulations for the unit commitment problems with ramping constraints. *Operations Research*.

[Beale and Tomlin, 1970] Beale, E. M. L. and Tomlin, J. A. (1970). Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables. *OR*, 69(447-454):99.

[Beck et al., 2023] Beck, Y., Ljubić, I., and Schmidt, M. (2023). A survey on bilevel optimization under uncertainty. *European Journal of Operational Research*.

[Beck and Schmidt, 2021] Beck, Y. and Schmidt, M. (2021). A gentle and incomplete introduction to bilevel optimization.

[Belotti et al., 2016] Belotti, P., Bonami, P., Fischetti, M., Lodi, A., Monaci, M., Nogales-Gómez, A., and Salvagnin, D. (2016). On handling indicator constraints in mixed integer programming. *Computational Optimization and Applications*, 65:545–566.

[Bertsimas et al., 2011] Bertsimas, D., Brown, D. B., and Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM review*, 53(3):464–501.

[Bestuzheva et al., 2023] Bestuzheva, K., Gleixner, A., and Achterberg, T. (2023). Efficient separation of RLT cuts for implicit and explicit bilinear products. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 14–28. Springer.

[Bezanson et al., 2017] Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98.

[Bialas and Karwan, 1978] Bialas, W. and Karwan, M. (1978). Multilevel linear programming. Technical report, State University of New York at Buffalo, NY.

[Bjørndal and Jörnsten, 2008] Bjørndal, M. and Jörnsten, K. (2008). Equilibrium prices supported by dual price functions in markets with non-convexities. *European Journal of Operational Research*, 190(3):768–789.

[Boiteux, 1960] Boiteux, M. (1960). Peak-load pricing. *The Journal of Business*, 33(2):157–179.

[Borghetti et al., 2003] Borghetti, A., Frangioni, A., Lacalandra, F., and Nucci, C. (2003). Lagrangian heuristics based on disaggregated bundle methods for hydrothermal unit commitment. *IEEE Transactions on Power Systems*, 18:313–323.

[Borrero et al., 2022] Borrero, J. S., Prokopyev, O. A., and Sauré, D. (2022). Learning in sequential bilevel linear programming. *INFORMS Journal on Optimization*, 4(2):174–199.

[Bouffard and Galiana, 2005] Bouffard, F. and Galiana, F. D. (2005). Generalized uplifts in pool-based electricity markets. In *Analysis, control and optimization of complex dynamic systems*, pages 193–214. Springer.

[Bowring and Tyler, 2019] Bowring, J. and Tyler, C. (2019). Calculating scarcity rents.

[Bracken and McGill, 1973] Bracken, J. and McGill, J. T. (1973). Mathematical programs with optimization problems in the constraints. *Operations research*, 21(1):37–44.

[Burer and Letchford, 2012] Burer, S. and Letchford, A. N. (2012). Non-convex mixed-integer nonlinear programming: A survey. *Surveys in Operations Research and Management Science*, 17:97–106.

[Byers and Eldridge, 2022] Byers, C. and Eldridge, B. (2022). Auction designs to increase incentive compatibility and reduce self-scheduling in electricity markets.

[Byers and Hug, 2023] Byers, C. and Hug, G. (2023). Long-run optimal pricing in electricity markets with non-convex costs. *European Journal of Operational Research*, 307(1):351–363.

[Bylling, 2018] Bylling, H. C. (2018). *Bilevel Optimization with Application in Energy*. PhD thesis, University of Copenhagen, Faculty of Science, Department of Mathematical Sciences.

[Candler and Norton, 1977] Candler, W. and Norton, R. D. (1977). *Multi-level programming*, volume 20. World Bank.

[Caramanis and Associates Inc., 2002] Caramanis, T. and Associates Inc. (2002). UPPO auction module user manual. Online; accessed 2023-02-01.

[Cazzola and Creatini, 2002] Cazzola, C. and Creatini, R. (2002). Equilibri non cooperativi e «caos» concorrenziale nel mercato elettrico italiano. *Rivista Internazionale di Scienze Sociali*, pages 431–466.

[Conejo et al., 2016] Conejo, A. J., Baringo, L., Kazempour, S. J., and Siddiqui, A. S. (2016). *Investment in electricity generation and transmission*, volume 119. Springer.

[Conejo et al., 2006] Conejo, A. J., Castillo, E., Minguez, R., and Garcia-Bertrand, R. (2006). *Decomposition techniques in mathematical programming: engineering and science applications*. Springer Science & Business Media.

[Cottle, 2012] Cottle, R. W. (2012). William Karush and the KKT theorem. *Doc. Math., Extra volume: Optimization stories*, pages 255–269.

[Coutu and White, 2014a] Coutu, R. and White, M. (2014a). Real-time price formation: Technical session nr. 1.

[Coutu and White, 2014b] Coutu, R. and White, M. (2014b). Real-time price formation: Technical session nr. 3.

[Cplex IBM, ILOG, 2022] Cplex IBM, ILOG (2022). V22.10.1: User's manual for CPLEX.

[Dempe, 2020] Dempe, S. (2020). Bilevel optimization: theory, algorithms, applications and a bibliography. *Bilevel Optimization: Advances and Next Challenges*, pages 581–672.

[Dempe et al., 2018] Dempe, S., Luo, G., and Franke, S. (2018). Pessimistic bilevel linear optimization. *Journal of Nepal Mathematical Society.*

[DeNegre and Ralphs, 2009] DeNegre, S. T. and Ralphs, T. K. (2009). A branch-and-cut algorithm for integer bilevel linear programs. In *Operations research and cyber-infrastructure*, pages 65–78. Springer.

[Deng, 1998] Deng, X. (1998). *Complexity Issues in Bilevel Linear Programming*, pages 149–164. Springer US, Boston, MA.

[Deng and Lv, 2020] Deng, X. and Lv, T. (2020). Power system planning with increasing variable renewable energy: A review of optimization models. *Journal of Cleaner Production*, 246:118962.

[Dias Garcia et al., 2023] Dias Garcia, J., Bodin, G., and Street, A. (2023). BilevelJuMP.jl: Modeling and Solving Bilevel Optimization Problems in Julia. *INFORMS Journal on Computing*, pages 1–9.

[Dongdong et al., 2022] Dongdong, G., Qi, H., Zizhuo, W., Jian, W., and Yinyu, Y. (2022). Cardinal Optimizer (COPT) user guide.

[Dvorkin et al., 2017] Dvorkin, Y., Fernández-Blanco, R., Kirschen, D. S., Pandžić, H., Watson, J.-P., and Silva-Monroy, C. A. (2017). Ensuring profitability of energy storage. *IEEE Transactions on Power Systems*, 32(1):611–623.

[Eldridge et al., 2018] Eldridge, B., O'Neill, R., and Hobbs, B. (2018). Pricing in day-ahead electricity markets with near-optimal unit commitment.

[Eldridge et al., 2020] Eldridge, B., O'Neill, R., and Hobbs, B. F. (2020). Near-optimal scheduling in day-ahead markets: Pricing models and payment redistribution bounds. *IEEE Transactions on Power Systems*, 35(3):1684–1694.

[European Commission, 2020] European Commission (2020). Italian implementation plan. Ref. Ares(2020)3317597.

[FICO, 2023] FICO (2023). FICO Xpress Optimization. Optimizer Reference Manual, January 2023.

[Fischetti and Lodi, 2010] Fischetti, M. and Lodi, A. (2010). Heuristics in mixed integer programming. *Wiley encyclopedia of operations research and management science.*

[Fortet, 1960] Fortet, R. (1960). L'algebre de boole et ses applications en recherche opérationnelle. *Trabajos de Estadistica*, 11:111–118.

[Fortuny-Amat and McCarl, 1981] Fortuny-Amat, J. and McCarl, B. (1981). A representation and economic interpretation of a two-level programming problem. *Journal of the operational Research Society*, 32:783–792.

[Frangioni, 2002] Frangioni, A. (2002). Generalized bundle methods. *SIAM Journal on Optimization*, 13(1):117–156.

[Frangioni, 2005] Frangioni, A. (2005). About Lagrangian Methods in Integer Optimization. *Annals of Operations Research*, 139(1):163–193.

[Frangioni and Lacalandra, 2022] Frangioni, A. and Lacalandra, F. (2022). A bilevel programming approach to price decoupling in pay-as-clear markets, with application to day-ahead electricity markets.

[Fraunholz et al., 2023] Fraunholz, C., Miskiw, K. K., Kraft, E., Fichtner, W., and Weber, C. (2023). On the role of risk aversion and market design in capacity expansion planning. *The Energy Journal*, 44(3):111–138.

[Frew et al., 2016] Frew, B. A., Milligan, M., Brinkman, G., Bloom, A., Clark, K., and Denholm, P. (2016). Revenue sufficiency and reliability in a zero marginal cost future. Technical report, National Renewable Energy Lab., Golden, CO, US.

[Galiana et al., 2003] Galiana, F., Motto, A., and Bouffard, F. (2003). Reconciling social welfare, agent profits, and consumer payments in electricity pools. *IEEE Transactions on Power Systems*, 18(2):452–459.

[García and Reitzes, 2007] García, J. A. and Reitzes, J. D. (2007). International perspectives on electricity market monitoring and market power mitigation. *Review of Network Economics*, 6(3).

[Gentile et al., 2017] Gentile, C., Morales-España, G., and Ramos, A. (2017). A tight MIP formulation of the unit commitment problem with start-up and shut-down constraints. *EURO Journal on Computational Optimization*, 5(1):177–201.

[GME, 2002] GME (2002). Determinazione dell'equilibrio del mercato del giorno prima dell'energia con prezzo uniforme per i consumatori e prezzi zonali per i generatori. Technical Report 5, GME.

[GME, 2016] GME (2016). Numero offerte multiple sul MGP, MI e MSD.

[GME, 2019] GME (2019). Vademecum della piattaforma dei conti energia a termine (PCE).

[Goyal and Richard, 2024] Goyal, A. and Richard, J.-P. P. (2024). Globally solving a class of bilevel programs with spatial price equilibrium constraints.

[Gribik et al., 2007] Gribik, P., Hogan, W. W., and Susan, P. (2007). Market clearing electricity prices and energy uplift. page 46.

[Guo et al., 2022] Guo, C., Bodur, M., and Papageorgiou, D. J. (2022). Generation expansion planning with revenue adequacy constraints. *Computers & Operations Research*, 142:105736.

[Gurobi Optimization, 2016] Gurobi Optimization (2016). Algorithms in gurobi. https://assets.gurobi.com/pdfs/user-events/2016-frankfurt/Die-Algorithmen.pdf.

[Gurobi Optimization, LLC, 2023] Gurobi Optimization, LLC (2023). Gurobi Optimizer Reference Manual.

[Hemmati et al., 2013] Hemmati, R., Hooshmand, R.-A., and Khodabakhshian, A. (2013). Comprehensive review of generation and transmission expansion planning. *IET Generation, Transmission & Distribution*, 7(9):955–964.

[Herrero et al., 2015] Herrero, I., Rodilla, P., and Batlle, C. (2015). Electricity market-clearing prices and investment incentives: The role of pricing rules. *Energy Economics*, 47:42–51.

[Herrero et al., 2020] Herrero, I., Rodilla, P., and Batlle, C. (2020). Evolving Bidding Formats and Pricing Schemes in USA and Europe Day-Ahead Electricity Markets. *Energies*, 13(19).

[Hua and Baldick, 2017] Hua, B. and Baldick, R. (2017). A convex primal formulation for convex hull pricing. *IEEE Transactions on Power Systems*, 32(5):3814–3823.

[Hytowitz et al., 2020] Hytowitz, R. B., Frew, B., Stephen, G., Ela, E., Singhal, N., Bloom, A., and Lau, J. (2020). Impacts of price formation efforts considering high renewable penetration levels and system resource adequacy targets. Technical report, National Renewable Energy Lab., Golden, CO, US.

[Hytowitz, 2018] Hytowitz, R. S. B. (2018). *Optimization modeling to address the impacts of electric power market design on operations and policy.* PhD thesis, The Johns Hopkins University.

[Karush, 1939] Karush, W. (1939). Minima of functions of several variables with inequalities as side constraints. *M. Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago.*

[Khun and Tucker, 1951] Khun, H. and Tucker, A. (1951). Non-linear programming, proceeding second Berkeley symposium Mathematical Statistic and probability (ed) Nyman, J. *Princeton University and Stanford University*, 481:492.

[Kleinert, 2021] Kleinert, T. (2021). *Algorithms for Mixed-Integer Bilevel Problems with Convex Followers.* PhD thesis.

[Kleinert et al., 2021a] Kleinert, T., Labbé, M., Ljubić, I., and Schmidt, M. (2021a). A survey on mixed-integer programming techniques in bilevel optimization. *EURO Journal on Computational Optimization*, 9:100007.

[Kleinert et al., 2021b] Kleinert, T., Labbé, M., Plein, F., and Schmidt, M. (2021b). Closing the gap in linear bilevel optimization: a new valid primal-dual inequality. *Optimization Letters*, 15:1027–1040.

[Kleinert et al., 2020] Kleinert, T., Labbé, M., Plein, F. a., and Schmidt, M. (2020). There's no free lunch: on the hardness of choosing a correct big-m in bilevel optimization. *Operations research*, 68(6):1716–1721.

[Koltsaklis and Dagoumas, 2018] Koltsaklis, N. E. and Dagoumas, A. S. (2018). State-of-the-art generation expansion planning: A review. *Applied Energy*, 230:563–589.

[Lalla-Ruiz and Voss, 2016] Lalla-Ruiz, E. and Voss, S. (2016). Improving solver performance through redundancy. *Journal of Systems Science and Systems Engineering*, 25:303–325.

[Li et al., 2022] Li, C., Conejo, A. J., Siirola, J. D., and Grossmann, I. E. (2022). On representative day selection for capacity expansion planning of power systems under extreme operating conditions. *International Journal of Electrical Power and Energy Systems*, 137:107697.

[Liberopoulos and Andrianesis, 2016] Liberopoulos, G. and Andrianesis, P. (2016). Critical review of pricing schemes in markets with non-convex costs. *Operations Research*, 64(1):17–31.

[Liberti, 2009] Liberti, L. (2009). Reformulations in mathematical programming: Definitions and systematics. *RAIRO-Operations Research*, 43(1):55–85.

[Liebensteiner and Naumann, 2022] Liebensteiner, M. and Naumann, F. (2022). Can carbon pricing counteract renewable energies' cannibalization problem? *Energy Economics*, 115:106345.

[Lubin et al., 2023] Lubin, M., Dowson, O., Dias Garcia, J., Huchette, J., Legat, B., and Vielma, J. P. (2023). JuMP 1.0: Recent improvements to a modeling language for mathematical optimization. *Mathematical Programming Computation*.

[MacDonald et al., 2023] MacDonald, J., Cappers, P., Callaway, D., and Kiliccote, S. (2023). Demand response providing ancillary services: A comparison of opportunities and challenges in us wholesale markets.

[Madani and Papavasiliou, 2022] Madani, M. and Papavasiliou, A. (2022). A note on a revenue adequate pricing scheme that minimizes make-whole payments. In *2022 18th International Conference on the European Energy Market (EEM)*, pages 1–6.

[Mays et al., 2021] Mays, J., Morton, David, P., and O'Neill, R. P. (2021). Investment effects of pricing schemes for non-convex markets. *European Journal of Operational Research*, 289(2):712–726.

[McCormick, 1976] McCormick, G. P. (1976). Computability of global solutions to factorable nonconvex programs: Part i—convex underestimating problems. *Mathematical programming*, 10(1):147–175.

[Meus et al., 2018] Meus, J., Poncelet, K., and Delarue, E. (2018). Applicability of a clustered unit commitment model in power system modeling. *IEEE Transactions on Power Systems*, 33(2):2195–2204.

[Micheli and Vespucci, 2020] Micheli, G. and Vespucci, M. T. (2020). *Mathematical models and solution algorithms for Generation and Transmission Expansion Planning with high shares of renewables*. PhD dissertation, University of Bergamo.

[Micheli et al., 2021] Micheli, G., Vespucci, M. T., Stabile, M., and Cortazzi, A. (2021). *Selecting and Initializing Representative Days for Generation and Transmission Expansion Planning with High Shares of Renewables*, pages 321–334. Springer International Publishing, Cham.

[Ministero dell'Ambiente e della Sicurezza Energetica, 2023] Ministero dell'Ambiente e della Sicurezza Energetica (2023). Consultazione pubblica schema di dm ferx.

[MOSEK ApS, 2023] MOSEK ApS (2023). *MOSEK Modeling Cookbook Release 3.3.0*.

[Motto and Galiana, 2002] Motto, A. and Galiana, F. (2002). Equilibrium of auction markets with unit commitment: the need for augmented pricing. *IEEE Transactions on Power Systems*, 17(3):798–805.

[Muñoz et al., 2013] Muñoz, M. P., Corchero, C., and Heredia, F.-J. (2013). Improving electricity market price forecasting with factor models for the optimal generation bid. *International Statistical Review*, 81(2):289–306.

[O'Neill et al., 2017] O'Neill, R. P., Castillo, A., Eldridge, B., and Hytowitz, R. B. (2017). Dual pricing algorithm in iso markets. *IEEE Transactions on Power Systems*, 32(4):3308–3310.

[O'Neill et al., 2005] O'Neill, R. P., Sotkiewicz, P. M., Hobbs, B. F., Rothkopf, M. H., and Stewart, W. R. (2005). Efficient market-clearing prices in markets with nonconvexities. *European Journal of Operational Research*, 164(1):269–285.

[Pablo Luna et al., 2021] Pablo Luna, J., Sagastizábal, C., and Silva, P. J. (2021). A discussion on electricity prices, or the two sides of the coin. *Philosophical Transactions of the Royal Society A*, 379(2202):20190428.

[Papavasiliou, 2021] Papavasiliou, A. (2021). Overview of eu capacity remuneration mechanisms. Technical report, Technical Report, URL https://www. rae. gr/wp-content/uploads/2021/05/Report ....

[Pereira et al., 2005] Pereira, M. V., Granville, S., Fampa, M. H., Dix, R., and Barroso, L. A. (2005). Strategic bidding under uncertainty: a binary expansion approach. *IEEE Transactions on Power Systems*, 20(1):180–188.

[Poncelet et al., 2020] Poncelet, K., Delarue, E., and D'haeseleer, W. (2020). Unit commitment constraints in long-term planning models: Relevance, pitfalls and the role of assumptions on flexibility. *Applied Energy*, 258(C).

[Pozo, 2022] Pozo, D. (2022). Linear battery models for power systems analysis. *Electric Power Systems Research*, 212:108565.

[Pozo, 2023] Pozo, D. (2023). Convex hull formulations for linear modeling of energy storage systems. *IEEE Transactions on Power Systems*, 38(6):5934–5936.

[Prol et al., 2020] Prol, J. L., Steininger, K. W., and Zilberman, D. (2020). The cannibalization effect of wind and solar in the california wholesale electricity market. *Energy Economics*, 85:104552.

[Quick and Carey, 2002] Quick, D. M. and Carey, J. M. (2002). Transmission capacity and market power: the effect on a dominant generation firm. *Energy Policy*, 30(8):699–708.

[Rajan et al., 2005] Rajan, D., Takriti, S., et al. (2005). Minimum up/down polytopes of the unit commitment problem with start-up costs. *IBM Res. Rep*, 23628:1–14.

[Roald et al., 2023] Roald, L. A., Pozo, D., Papavasiliou, A., Molzahn, D. K., Kazempour, J., and Conejo, A. (2023). Power systems optimization under uncertainty: A review of methods and applications. *Electric Power Systems Research*, 214:108725.

[Rodilla et al., 2013] Rodilla, P., Cerisola, S., and Batlle, C. (2013). Modeling the major overhaul cost of gas-fired plants in the unit commitment problem. *IEEE Transactions on Power Systems*, 29(3):1001–1011.

[Ruiz et al., 2012] Ruiz, C., Conejo, A. J., and Gabriel, S. A. (2012). Pricing non-convexities in an electricity pool. *IEEE Transactions on Power Systems*, 27(3):1334–1342.

[Ruiz and Grossmann, 2011] Ruiz, J. P. and Grossmann, I. E. (2011). Using redundancy to strengthen the relaxation for the global optimization of minlp problems. *Computers & Chemical Engineering*, 35(12):2729–2740.

[Sahinidis, 2019] Sahinidis, N. V. (2019). Mixed-integer nonlinear programming 2018.

[Schiro et al., 2016] Schiro, D. A., Zheng, T., Zhao, F., and Litvinov, E. (2016). Convex hull pricing in electricity markets: Formulation, analysis, and implementation challenges. *IEEE Transactions on Power Systems*, 31(5):4068–4075.

[Sherali and Adams, 1990] Sherali, H. D. and Adams, W. P. (1990). A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3(3):411–430.

[Sherali and Adams, 1999] Sherali, H. D. and Adams, W. P. (1999). A reformulation-linearization technique for solving discrete and continuous nonconvex problems.

[Siddiqui and Gabriel, 2013] Siddiqui, S. and Gabriel, S. A. (2013). An sos1-based approach for solving mpecs with a natural gas market application. *Networks and Spatial Economics*, 13:205–227.

[Sinha et al., 2017] Sinha, A., Malo, P., and Deb, K. (2017). A review on bilevel optimization: From classical to evolutionary approaches and applications. *IEEE Transactions on Evolutionary Computation*, 22(2):276–295.

[Stackelberg, 1952] Stackelberg, H. (1952). The theory of the market economy, oxford university, press.

[Stevens and Papavasiliou, 2022] Stevens, N. and Papavasiliou, A. (2022). Application of the level method for computing locational convex hull prices. *IEEE Transactions on Power Systems*, 37(5):3958–3968.

[Taylor, 2018] Taylor, J. (2018). *Convex Optimization of Power Systems*.

[Terna, 2021] Terna (2021). The new electricity market zones: what you need to know.

[Tibshirani et al., 2001] Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423.

[Ueckerdt et al., 2013] Ueckerdt, F., Hirth, L., Luderer, G., and Edenhofer, O. (2013). System lcoe: What are the costs of variable renewables? *Energy*, 63:61–75.

[Van Ackooij et al., 2018] Van Ackooij, W., Danti Lopez, I., Frangioni, A., Lacalandra, F., and Tahanan, M. (2018). Large-scale unit commitment under uncertainty: an updated literature survey. *Annals of Operations Research*, 271(1):11–85.

[Van Vyve, 2011] Van Vyve, M. (2011). Linear prices for non-convex electricity markets: models and algorithms. Technical report, CORE Leuven, Belgium.

[Vanderbeck and Savelsbergh, 2006] Vanderbeck, F. and Savelsbergh, M. W. (2006). A generic view of Dantzig–Wolfe decomposition in mixed integer programming. *Operations Research Letters*, 34(3):296–306.

[Vespucci et al., 2013] Vespucci, M. T., Innorta, M., and Cervigni, G. (2013). A mixed integer linear programming model of a zonal electricity market with a dominant producer. *Energy Economics*, 35:35–41.

[Vielma, 2015] Vielma, J. P. (2015). Mixed integer linear programming formulation techniques. *Siam Review*, 57(1):3–57.

[Von Stackelberg, 1934] Von Stackelberg, H. (1934). Marktform and gleichgewicht springer-verlag.

[Wang et al., 2016] Wang, C., Luh, P. B., Gribik, P., Peng, T., and Zhang, L. (2016). Commitment cost allocation of fast-start units for approximate extended locational marginal prices. *IEEE Transactions on Power Systems*, 31(6):4176–4184.

[Wogrin et al., 2011] Wogrin, S., Centeno, E., and Barquin, J. (2011). Generation capacity expansion in liberalized electricity markets: A stochastic mpec approach. *IEEE Transactions on Power Systems*, 26(4):2526–2532.

[Wuijts et al., 2023] Wuijts, R. H., van den Akker, M., and van den Broek, M. (2023). New efficient ADMM algorithm for the unit commitment problem.

[Yıldız and Vielma, 2013] Yıldız, S. and Vielma, J. P. (2013). Incremental and encoding formulations for mixed integer programming. *Operations Research Letters*, 41(6):654–658.

[Zare et al., 2019] Zare, M. H., Borrero, J. S., Zeng, B., and Prokopyev, O. A. (2019). A note on linearized reformulations for a class of bilevel linear integer problems. *Annals of Operations Research*, 272:99–117.

Appendix A

The CSC of the lower level problem

Starting from the *primal* of the LL (2) problem defined for each $t \in T$ and provided again for convenience, the corresponding *dual* is subsequently derived as (A.2).

$$\min_{Q,F_{l,t}} LLP = \sum_{z \in Z} \left(P^{V}\, Q^{V}_{z,t} + P^{WI}\, Q^{WI}_{z,t} + P^{WO}\, Q^{WO}_{z,t} \right) + \sum_{k \in K} P^{K}_{k}\, Q^{K}_{k,t} \tag{5.42a}$$

$$Q^{V}_{z,t} + Q^{WI}_{z,t} + Q^{WO}_{z,t} + \sum_{k \in K_z} Q^{K}_{k,t} + B^{D}_{z,t}$$

$$+ \sum_{l \in BS_z} F_{l,t} \geq D_{z,t} + \sum_{l \in FS_z} F_{l,t} + B^{C}_{z,t} \qquad z \in Z \qquad (\lambda_{z,t} \geq 0) \tag{5.62b}$$

$$0 \leq Q^{V}_{z,t} \leq cf^{V}_{z,t}\, \overline{Q}^{V}\, (n^{V}_{z,0} + n^{V}_{z}) \qquad z \in Z \qquad (\mu^{V}_{z,t} \geq 0) \tag{5.42c}$$

$$0 \leq Q^{WI}_{z,t} \leq cf^{WI}_{z,t}\, \overline{Q}^{WI}\, (n^{WI}_{z,0} + n^{WI}_{z}) \qquad z \in Z \qquad (\mu^{WI}_{z,t} \geq 0) \tag{5.42d}$$

$$0 \leq Q^{WO}_{z,t} \leq cf^{WO}_{z,t}\, \overline{Q}^{WO}\, (n^{WO}_{z,0} + n^{WO}_{z}) \qquad z \in Z \qquad (\mu^{WO}_{z,t} \geq 0) \tag{5.42e}$$

$$\underline{Q}^{K}_{k}\, \gamma_{k,t} \leq Q^{K}_{k,t} \leq \overline{Q}^{K}_{k}\, \gamma_{k,t} \qquad k \in K \qquad (\underline{\mu}^{K}_{k,t} \geq 0, \overline{\mu}^{K}_{k,t} \geq 0) \tag{5.42f}$$

$$\underline{F}_{l} \leq F_{l,t} \leq \overline{F}_{l} \qquad l \in L \qquad (\underline{\upsilon}_{l,t} \geq 0, \overline{\upsilon}_{l,t} \geq 0) \tag{5.42g}$$

$$\max_{\substack{\lambda_{z,t},\\ \underline{\mu}^K_{k,t},\overline{\mu}^K_{k,t},\\ \mu^V_{z,t},\mu^{WI}_{z,t},\mu^{WO}_{z,t},\\ \underline{\upsilon}_{l,t},\overline{\upsilon}_{l,t}}} \quad LLD \tag{A.2a}$$

$$\begin{aligned} LLD = &\sum_{z\in Z} \left(D_{z,t} - B^D_{z,t} + B^C_{z,t}\right)\ \lambda_{z,t} - \sum_{z\in Z}\left[cf^V_{z,t}\ \overline{Q}^V\ (n^V_{z,0} + n^V_z)\ \mu^V_{z,t}\right] \\ &- \sum_{z\in Z}\left[cf^{WI}_{z,t}\ \overline{Q}^{WI}\ (n^{WI}_{z,0} + n^{WI}_z)\ \mu^{WI}_{z,t}\right] - \sum_{z\in Z}\left[cf^{WO}_{z,t}\ \overline{Q}^{WO}\ (n^{WO}_{z,0} + n^{WO}_z)\ \mu^{WO}_{z,t}\right] \\ &+ \sum_{k\in K}\left(\underline{Q}^K_k\ \gamma_{k,t}\ \underline{\mu}^K_{k,t} - \overline{Q}^K_k\ \gamma_{k,t}\ \overline{\mu}^K_{k,t}\right) + \sum_{l\in L}\left(\underline{F}_l\ \underline{\upsilon}_{l,t} - \overline{F}_l\ \overline{\upsilon}_{l,t}\right) \end{aligned} \tag{A.2b}$$

$$P^V - \lambda_{z,t} + \mu^V_{z,t} \geq 0 \qquad z\in Z \qquad (Q^V_{z,t}\geq 0) \tag{A.2c}$$

$$P^{WI} - \lambda_{z,t} + \mu^{WI}_{z,t} \geq 0 \qquad z\in Z \qquad (Q^{WI}_{z,t}\geq 0) \tag{A.2d}$$

$$P^{WO} - \lambda_{z,t} + \mu^{WO}_{z,t} \geq 0 \qquad z\in Z \qquad (Q^{WO}_{z,t}\geq 0) \tag{A.2e}$$

$$P^K_k - \lambda_{z,t} - \underline{\mu}^K_{k,t} + \overline{\mu}^K_{k,t} \qquad k\in K \qquad (Q^K_{k,t}\geq 0) \tag{A.2f}$$

$$\lambda_{z^{tail}_l} - \lambda_{z^{head}_l} - \underline{\upsilon}_{l,t} + \overline{\upsilon}_{l,t} = 0 \qquad l\in L \qquad (F_{l,t}\ free) \tag{A.2g}$$

The *Product formulation* of the CSC of LL (2) for each $t \in T$ is thus defined as (6.1).

$$\left(Q^V_{z,t}+Q^{WI}_{z,t}+Q^{WO}_{z,t}+\textstyle\sum_{k\in K_z} Q^K_{k,t}+\sum_{l\in BS_z} F_{l,t}-\sum_{l\in FS_z} F_{l,t}-D_{z,t}+B^D_{z,t}-B^C_{z,t}\right) \perp \lambda_{z,t} \qquad z\in Z \tag{6.1a}$$

$$\left(cf^V_{z,t}\cdot\overline{Q}^V\cdot(n^V_{z,0}+n^V_z) - Q^V_{z,t}\right) \perp \mu^V_{z,t} \qquad z\in Z \tag{6.1d}$$

$$\left(cf^{WI}_{z,t}\cdot\overline{Q}^{WI}\cdot(n^{WI}_{z,0}+n^{WI}_z) - Q^{WI}_{z,t}\right) \perp \mu^{WI}_{z,t} \qquad z\in Z \tag{6.1e}$$

$$\left(cf^{WO}_{z,t}\cdot\overline{Q}^{WO}\cdot(n^{WO}_{z,0}+n^{WO}_z) - Q^{WO}_{z,t}\right) \perp \mu^{WO}_{z,t} \qquad z\in Z \tag{6.1f}$$

$$\left(Q^K_{k,t} - \underline{Q}^K_k\cdot\gamma_{k,t}\right) \perp \underline{\mu}^K_{k,t} \qquad k\in K \tag{6.1g}$$

$$\left(\overline{Q}^K_k\cdot\gamma_{k,t} - Q^K_{k,t}\right) \perp \overline{\mu}^K_{k,t} \qquad k\in K \tag{6.1h}$$

$$\left(F_{l,t} - \underline{F}_l\right) \perp \underline{\upsilon}_{l,t} \qquad l\in L \tag{6.1b}$$

$$\left(\overline{F}_l - F_{l,t}\right) \perp \overline{\upsilon}_{l,t} \qquad l\in L \tag{6.1c}$$

$$\left(P_k^K - \lambda_{z,t} - \underline{\mu}_{k,t}^K + \overline{\mu}_{k,t}^K\right) \perp Q_{k,t}^K \qquad k \in K_z, z \in Z \quad (6.1l)$$

$$\left(P^V - \lambda_{z,t} + \mu_{z,t}^V\right) \perp Q_{z,t}^V \qquad z \in Z \quad (6.1i)$$

$$\left(P^{WI} - \lambda_{z,t} + \mu_{z,t}^{WI}\right) \perp Q_{z,t}^{WI} \qquad z \in Z \quad (6.1j)$$

$$\left(P^{WO} - \lambda_{z,t} + \mu_{z,t}^{WO}\right) \perp Q_{z,t}^{WO} \qquad z \in Z \quad (6.1k)$$

The *single constraint* aggregating all the CSC conditions for each $t \in T$ is defined as (A.3).

$$\begin{aligned}
&\sum_{z\in Z}\left(Q_{z,t}^V + Q_{z,t}^{WI} + Q_{z,t}^{WO} + \sum_{k\in K_z} Q_{k,t}^K + \sum_{l\in BS_z} F_{l,t} - \sum_{l\in FS_z} F_{l,t} - D_{z,t} + B_{z,t}^D - B_{z,t}^C\right)\cdot \lambda_{z,t}\\
&+ \sum_{z\in Z}\left(cf_{z,t}^V \cdot \overline{Q}^V \cdot (n_{z,0}^V + n_z^V) - Q_{z,t}^V\right)\cdot \mu_{z,t}^V + \sum_{z\in Z}\left(cf_{z,t}^{WI}\cdot \overline{Q}^{WI}\cdot (n_{z,0}^{WI} + n_z^{WI}) - Q_{z,t}^{WI}\right)\cdot \mu_{z,t}^{WI}\\
&+ \sum_{z\in Z}\left(cf_{z,t}^{WO}\cdot \overline{Q}^{WO}\cdot (n_{z,0}^{WO} + n_z^{WO}) - Q_{z,t}^{WO}\right)\cdot \mu_{z,t}^{WO} + \sum_{k\in K}\left(Q_{k,t}^K - \underline{Q}_k^K \cdot \gamma_{k,t}\right)\cdot \underline{\mu}_{k,t}^K\\
&+ \sum_{k\in K}\left(\overline{Q}_k^K \cdot \gamma_{k,t} - Q_{k,t}^K\right)\cdot \overline{\mu}_{k,t}^K + \sum_{l\in L}(F_{l,t} - \underline{F}_l)\cdot \underline{v}_{l,t} + \sum_{l\in L}\left(\overline{F}_l - F_{l,t}\right)\cdot \overline{v}_{l,t}\\
&+ \sum_{z\in Z}\sum_{k\in K_z}\left(P_k^K - \lambda_{z,t} - \underline{\mu}_{k,t}^K + \overline{\mu}_{k,t}^K\right)\cdot Q_{k,t}^K + \sum_{z\in Z}\left(P^V - \lambda_{z,t} + \mu_{z,t}^V\right)\cdot Q_{z,t}^V\\
&+ \sum_{z\in Z}\left(P^{WI} - \lambda_{z,t} + \mu_{z,t}^{WI}\right)\cdot Q_{z,t}^{WI} + \sum_{z\in Z}\left(P^{WO} - \lambda_{z,t} + \mu_{z,t}^{WO}\right)\cdot Q_{z,t}^{WO} = 0 \qquad (A.3)
\end{aligned}$$

The *SOS1 reformulation* of the CSC for each $t \in T$ is defined as (A.4).

For each $z \in Z$:

$$s_{z,t,1}^{CC1} = Q_{z,t}^V + Q_{z,t}^{WI} + Q_{z,t}^{WO} + \sum_{k\in K_z} Q_{k,t}^K + \sum_{l\in BS_z} F_{l,t} - \sum_{l\in FS_z} F_{l,t} - D_{z,t} + B_{z,t}^D - B_{z,t}^C$$

$$s_{z,t,2}^{CC1} = \lambda_{z,t} \qquad (A.4a)$$

$$s_{z,t,1}^{CC2} = cf_{z,t}^V \cdot \overline{Q}^V \cdot (n_{z,0}^V + n_z^V) - Q_{z,t}^V \qquad s_{z,t,2}^{CC2} = \mu_{z,t}^V \quad (A.4b)$$

$$s_{z,t,1}^{CC3} = cf_{z,t}^{WI} \cdot \overline{Q}^{WI} \cdot (n_{z,0}^{WI} + n_z^{WI}) - Q_{z,t}^{WI} \qquad s_{z,t,2}^{CC3} = \mu_{z,t}^{WI} \quad (A.4c)$$

$$s_{z,t,1}^{CC4} = cf_{z,t}^{WO} \cdot \overline{Q}^{WO} \cdot (n_{z,0}^{WO} + n_z^{WO}) - Q_{z,t}^{WO} \qquad s_{z,t,2}^{CC4} = \mu_{z,t}^{WO} \quad (A.4d)$$

$$\{s_{z,t,1}^{CC1}, s_{z,t,2}^{CC1}\}, \{s_{z,t,1}^{CC2}, s_{z,t,2}^{CC2}\}, \{s_{z,t,1}^{CC3}, s_{z,t,2}^{CC3}\}, \{s_{z,t,1}^{CC4}, s_{z,t,2}^{CC4}\} \in SOS1 \qquad (A.4e)$$

For each $k \in K$:

$$s^{CC5}_{k,t,1} = Q^K_{k,t} - \underline{Q}^K_k \cdot \gamma_{k,t} \qquad s^{CC5}_{k,t,2} = \underline{\mu}^K_{k,t} \quad \text{(A.4f)}$$

$$s^{CC6}_{k,t,1} = \overline{Q}^K_k \; \gamma_{k,t} - Q^K_{k,t} \qquad s^{CC6}_{k,t,2} = \overline{\mu}^K_{k,t} \quad \text{(A.4g)}$$

$$\{s^{CC5}_{k,t,1}, s^{CC5}_{k,t,2}\}, \{s^{CC6}_{k,t,1}, s^{CC6}_{k,t,2}\} \in SOS1 \quad \text{(A.4h)}$$

For each $l \in L$:

$$s^{CC7}_{l,t,1} = F_{l,t} - \underline{F}_l \qquad s^{CC7}_{l,t,2} = \underline{\upsilon}_{l,t} \quad \text{(A.4i)}$$

$$s^{CC8}_{l,t,1} = \overline{F}_l - F_{l,t} \qquad s^{CC8}_{l,t,2} = \overline{\upsilon}_{l,t} \quad \text{(A.4j)}$$

$$\{s^{CC7}_{l,t,1}, s^{CC7}_{l,t,2}\}, \{s^{CC8}_{l,t,1}, s^{CC8}_{l,t,2}\} \in SOS1 \quad \text{(A.4k)}$$

For each $k \in K$:

$$s^{CC9}_{k,t,1} = P^K_k - \lambda_{z,t} - \underline{\mu}^K_{k,t} + \overline{\mu}^K_{k,t} \qquad s^{CC9}_{k,t,2} = Q^K_{k,t} \quad \text{(A.4l)}$$

$$\{s^{CC9}_{k,t,1}, s^{CC9}_{k,t,2}\} \in SOS1 \quad \text{(A.4m)}$$

For each $z \in Z$:

$$s^{CC10}_{z,t,1} = P^V - \lambda_{z,t} + \mu^V_{z,t} \qquad s^{CC10}_{z,t,2} = Q^V_{z,t} \quad \text{(A.4n)}$$

$$s^{CC11}_{z,t,1} = P^{WI} - \lambda_{z,t} + \mu^{WI}_{z,t} \qquad s^{CC11}_{z,t,2} = Q^{WI}_{z,t} \quad \text{(A.4o)}$$

$$s^{CC12}_{z,t,1} = P^{WO} - \lambda_{z,t} + \mu^{WO}_{z,t} \qquad s^{CC12}_{z,t,2} = Q^{WO}_{z,t} \quad \text{(A.4p)}$$

$$\{s^{CC10}_{z,t,1}, s^{CC10}_{z,t,2}\}, \{s^{CC11}_{z,t,1}, s^{CC11}_{z,t,2}\}, \{s^{CC12}_{z,t,1}, s^{CC12}_{z,t,2}\} \in SOS1 \quad \text{(A.4q)}$$

The *Big-M reformulation* of the CSC for each $t \in T$ is defined as (A.5).

For each $z \in Z$, $t \in T$:

$$Q^V_{z,t} + Q^{WI}_{z,t} + Q^{WO}_{z,t} + \sum_{k \in K_z} Q^K_{k,t} + \sum_{l \in BS_z} F_{l,t} - \sum_{l \in FS_z} F_{l,t} - D_{z,t} + B^D_{z,t} - B^C_{z,t} \leq M^{cp} \cdot (1 - b^{CC1}_{z,t})$$

$$\lambda_{z,t} \leq M^{vd} \cdot b^{CC1}_{z,t} \quad \text{(A.5a)}$$

$$cf^{V}_{z,t} \cdot \overline{Q}^{V} \cdot (n^{V}_{z,0} + n^{V}_{z}) - Q^{V}_{z,t} \leq M^{cp} \cdot (1 - b^{CC2}_{z,t}) \qquad \mu^{V}_{z,t} \leq M^{vd} \cdot b^{CC2}_{z,t} \tag{A.5b}$$

$$cf^{WI}_{z,t} \cdot \overline{Q}^{WI} \cdot (n^{WI}_{z,0} + n^{WI}_{z}) - Q^{WI}_{z,t} \leq M^{cp} \cdot (1 - b^{CC3}_{z,t}) \qquad \mu^{WI}_{z,t} \leq M^{vd} \cdot b^{CC3}_{z,t} \tag{A.5c}$$

$$cf^{WO}_{z,t} \cdot \overline{Q}^{WO} \cdot (n^{WO}_{z,0} + n^{WO}_{z}) - Q^{WO}_{z,t} \leq M^{cp} \cdot (1 - b^{CC4}_{z,t}) \qquad \mu^{WO}_{z,t} \leq M^{vd} \cdot b^{CC4}_{z,t} \tag{A.5d}$$

$$b^{CC1}_{z,t}, b^{CC2}_{z,t}, b^{CC3}_{z,t}, b^{CC4}_{z,t} \in \{0,1\} \tag{A.5e}$$

For each $k \in K$, $t \in T$:

$$Q^{K}_{k,t} - \underline{Q}^{K}_{k} \cdot \gamma_{k,t} \leq M^{cp} \cdot (1 - b^{CC5}_{k,t}) \qquad \underline{\mu}^{K}_{k,t} \leq M^{vd} \cdot b^{CC5}_{k,t} \tag{A.5f}$$

$$\overline{Q}^{K}_{k} \cdot \gamma_{k,t} - Q^{K}_{k,t} \leq M^{cp} \cdot (1 - b^{CC6}_{k,t}) \qquad \overline{\mu}^{K}_{k,t} \leq M^{vd} \cdot b^{CC6}_{k,t} \tag{A.5g}$$

$$b^{CC5}_{k,t}, b^{CC6}_{k,t} \in \{0,1\} \tag{A.5h}$$

For each $l \in L$, $t \in T$:

$$F_{l,t} - \underline{F}_{l} \leq M^{cp} \cdot (1 - b^{CC7}_{l,t}) \qquad \underline{\upsilon}_{l,t} \leq M^{vd} \cdot b^{CC7}_{l,t} \tag{A.5i}$$

$$\overline{F}_{l} - F_{l,t} \leq M^{cp} \cdot (1 - b^{CC8}_{l,t}) \qquad \overline{\upsilon}_{l,t} \leq M^{vd} \cdot b^{CC8}_{l,t} \tag{A.5j}$$

$$b^{CC7}_{l,t}, b^{CC8}_{l,t} \in \{0,1\} \tag{A.5k}$$

For each $k \in K$, $t \in T$:

$$P^{K}_{k} - \lambda_{z,t} - \underline{\mu}^{K}_{k,t} + \overline{\mu}^{K}_{k,t} \leq M^{cd} \cdot (1 - b^{CC9}_{k,t}) \qquad Q^{K}_{k,t} \leq M^{vp} \cdot b^{CC9}_{k,t} \tag{A.5l}$$

$$b^{CC9}_{k,t} \in \{0,1\} \tag{A.5m}$$

For each $z \in Z$, $t \in T$:

$$P^{V} - \lambda_{z,t} + \mu^{V}_{z,t} \leq M^{cd} \cdot (1 - b^{CC10}_{z,t}) \qquad Q^{V}_{z,t} \leq M^{vp} \cdot b^{CC10}_{z,t} \tag{A.5n}$$

$$P^{WI} - \lambda_{z,t} + \mu^{WI}_{z,t} \leq M^{cd} \cdot (1 - b^{CC11}_{z,t}) \qquad Q^{WI}_{z,t} \leq M^{vp} \cdot b^{CC11}_{z,t} \tag{A.5o}$$

$$P^{WO} - \lambda_{z,t} + \mu^{WO}_{z,t} \leq M^{cd} \cdot (1 - b^{CC12}_{z,t}) \qquad Q^{WO}_{z,t} \leq M^{vp} \cdot b^{CC12}_{z,t} \tag{A.5p}$$

$$b^{CC10}_{z,t}, b^{CC11}_{z,t}, b^{CC12}_{z,t} \in \{0,1\} \tag{A.5q}$$

The *SD reformulation* of the CSC for each $t \in T$ is defined as

$$LLP = LLD \tag{6.3}$$

The primal objective functions of LL (1) and LL (2) are identical while the dual objective functions differ as in LL (1) the charge is not considered in the balance constraint:

$$LLP = \sum_{z \in Z} \left(P^V Q^V_{z,t} + P^{WI} Q^{WI}_{z,t} + P^{WO} Q^{WO}_{z,t} \right) + \sum_{k \in K} P^K_k Q^K_{k,t}$$

$$\begin{aligned} LLD_1 = & \sum_{z \in Z} \left(D_{z,t} - B^D_{z,t} \right) \lambda_{z,t} - \sum_{z \in Z} cf^V_{z,t} \, \overline{Q}^V \, (n^V_{z,0} + n^V_z) \, \mu^V_{z,t} \\ & - \sum_{z \in Z} \left[cf^{WI}_{z,t} \, \overline{Q}^{WI} \, (n^{WI}_{z,0} + n^{WI}_z) \, \mu^{WI}_{z,t} + cf^{WO}_{z,t} \, \overline{Q}^{WO} \, (n^{WO}_{z,0} + n^{WO}_z) \, \mu^{WO}_{z,t} \right] \\ & + \sum_{k \in K} \left(\underline{Q}^K_k \, \gamma_{k,t} \, \underline{\mu}^K_{k,t} - \overline{Q}^K_k \, \gamma_{k,t} \, \overline{\mu}^K_{k,t} \right) + \sum_{l \in L} \left(\underline{F}_l \, \underline{v}_{l,t} - \overline{F}_l \, \overline{v}_{l,t} \right) \end{aligned}$$

$$\begin{aligned} LLD_2 = & \sum_{z \in Z} \left(D_{z,t} - B^D_{z,t} + B^C_{z,t} \right) \lambda_{z,t} - \sum_{z \in Z} cf^V_{z,t} \, \overline{Q}^V \, (n^V_{z,0} + n^V_z) \mu^V_{z,t} \\ & - \sum_{z \in Z} \left[cf^{WI}_{z,t} \, \overline{Q}^{WI} \, (n^{WI}_{z,0} + n^{WI}_z) \, \mu^{WI}_{z,t} + cf^{WO}_{z,t} \, \overline{Q}^{WO} \, (n^{WO}_{z,0} + n^{WO}_z) \, \mu^{WO}_{z,t} \right] \\ & + \sum_{k \in K} \left(\underline{Q}^K_k \, \gamma_{k,t} \, \underline{\mu}^K_{k,t} - \overline{Q}^K_k \, \gamma_{k,t} \, \overline{\mu}^K_{k,t} \right) + \sum_{l \in L} \left(\underline{F}_l \, \underline{v}_{l,t} - \overline{F}_l \, \overline{v}_{l,t} \right) \end{aligned}$$

Considering LLD_2, the SD equality is thus defined as (A.6) or equivalently (A.7).

$$\begin{aligned} & \sum_{z \in Z} (P^V Q^V_{z,t} + P^{WI} Q^{WI}_{z,t} + P^{WO} Q^{WO}_{z,t}) + \sum_{k \in K} P^K_k Q^K_{k,t} = \sum_{z \in Z} (D_{z,t} - B^D_{z,t} + B^C_{z,t}) \, \lambda_{z,t} \\ & - \sum_{z \in Z} \left[cf^V_{z,t} \, \overline{Q}^V \, (n^V_{z,0} + n^V_z) \, \mu^V_{z,t} + cf^{WI}_{z,t} \, \overline{Q}^{WI} \, (n^{WI}_{z,0} + n^{WI}_z) \, \mu^{WI}_{z,t} + cf^{WO}_{z,t} \, \overline{Q}^{WO} \, (n^{WO}_{z,0} + n^{WO}_z) \, \mu^{WO}_{z,t} \right] \\ & + \sum_{k \in K} \left(\underline{Q}^K_k \, \gamma_{k,t} \, \underline{\mu}^K_{k,t} - \overline{Q}^K_k \, \gamma_{k,t} \, \overline{\mu}^K_{k,t} \right) + \sum_{l \in L} (\underline{F}_l \, \underline{v}_{l,t} - \overline{F}_l \, \overline{v}_{l,t}) \end{aligned} \tag{A.6}$$

$$\begin{aligned} & \sum_{z \in Z} (D_{z,t} - B^D_{z,t} + B^C_{z,t}) \, \lambda_{z,t} = \sum_{z \in Z} (P^V Q^V_{z,t} + P^{WI} Q^{WI}_{z,t} + P^{WO} Q^{WO}_{z,t}) + \sum_{k \in K} P^K_k Q^K_{k,t} \\ & + \sum_{z \in Z} \left[cf^V_{z,t} \, \overline{Q}^V \, (n^V_{z,0} + n^V_z) \, \mu^V_{z,t} + cf^{WI}_{z,t} \, \overline{Q}^{WI} \, (n^{WI}_{z,0} + n^{WI}_z) \, \mu^{WI}_{z,t} + cf^{WO}_{z,t} \, \overline{Q}^{WO} \, (n^{WO}_{z,0} + n^{WO}_z) \, \mu^{WO}_{z,t} \right] \\ & - \sum_{k \in K} \left(\underline{Q}^K_k \, \gamma_{k,t} \, \underline{\mu}^K_{k,t} - \overline{Q}^K_k \, \gamma_{k,t} \, \overline{\mu}^K_{k,t} \right) - \sum_{l \in L} (\underline{F}_l \, \underline{v}_{l,t} - \overline{F}_l \, \overline{v}_{l,t}) \end{aligned} \tag{A.7}$$

Therefore, the objective function (5.45) of the GEP-RA (2) can be reformulated by replacing the left-hand terms of (A.7) with the right-hand terms, albeit with the addition of new bilinear terms.